VOLUME XXX APRIL, 1940 PART 4

THE RELATION OF COGNITION AND PHANTASY IN CHILDREN.

BY M. D. VERNON

From the Cambridge Psychological Laboratory

PART I

I. *Introduction* (pp. 273–275).

II. *Method of experiment* (pp. 275–278).

III. *Normal cognition and phantasy* (pp. 278–294).

(1) *Perception and immediate recall of meaningful pictures* (pp. 278–283).

(2) *Perception of meaningless designs* (pp. 283–287).

(3) *Free construction* (pp. 288–294).

References (p. 294).

I. INTRODUCTION

IT has been shown by Ruth Griffiths, in her book *Imagination in Early Childhood* (5), that the little child, of five years or thereabouts, is not able to think coherently and logically for more than a very short space of time. To deal with the many complicated emotional problems which he encounters, he uses the method of phantasy; that is to say, these problems and the people and situations they involve are represented in the child's mind symbolically, in a series of images. The phantasy of the little child is to a great extent divorced from reality, in the same way as is the adult dream; the situations and activities that occur have a 'magic' quality—they are not such as could happen in ordinary life, although they symbolize something that could or has happened. But Ruth Griffiths maintains that if the child is given suitable opportunities for expressing overtly and working out his phantasied emotional problems, in stories, drawings, etc., they will become progressively 'real'-ized. Also the child will become more fully aware of them, and they will be less dissociated from the rest of the personality, particularly from his cognitive activities.

Children, and particularly slightly older children, spontaneously express their phantasied wishes in imaginative play. Some plays of this kind have been fully examined and described by Margaret Lowenfeld in *Play in Childhood* (7). Among these children the 'magic' and symbolic elements of the phantasy were still very strong. A rather more 'real'-ized

type of imaginative play is observed in the 'robber' and 'desert island' play of small boys, and the 'home-making', 'doll-nursing' and 'school-teaching' play of little girls.

The typical childish phantasy is of the kind termed by Jung 'passive'[1] phantasy (6); it represents an irruption into consciousness of some unconscious mental process, the attitude towards it of the conscious individual being almost wholly passive. Thus it is 'self-active', and in the nature of a psychic 'automatism'. It seems probable that the adolescent day-dream or 'castle-in-the-air' is of the same nature, although its 'manifest content' is much more highly socialized than that of the childish phantasy. That is to say, the underlying wishes and desires are expressed in a form which is more or less completely 'real'-ized and non-magical, although their fundamental nature is the same as that of the childish wish.

But in the older child and the adult there appears in imaginative activity the 'active phantasy' of Jung (6). Here the conscious mind actively participates in selecting elements perceived intuitively in the unconscious mental contents, developing, rearranging and synthesizing them into the imaginative construction. Conscious cognitive activity necessarily plays a large part in such processes; and more especially in the imaginative interpretation of the constructions of others. Such imaginative interpretations can frequently be obtained from older children and adults if they are shown a picture and asked: "Tell me about it." But their responses will not as a rule be based upon simple perception and description of what is portrayed in the picture. Perception will be followed by imaginative reconstruction of situations and emotional states which the individual may never have experienced except in phantasy.

It is quite clear, however, from the work of Binet (2), Ségers (8) and others, that young children are incapable of this type of 'interpretation' (the term is Binet's); the actual age at which it appears is a matter of some disagreement, and will be discussed more fully below. But at some time or other in the development of the normal individual, unconscious mental contents cease to appear in waking life in the form of 'self-active' phantasy, though still manifesting themselves in their original symbolic and 'magic' form in dreams. (They may also appear in the adolescent 'castle-in-the-air', though possibly this is to some extent a pathological form.) Their normal mode of expression in waking life is in imaginative

[1] This term does not seem to be a good one, since the 'passive' phantasy is characteristically 'self-active'. Therefore as far as possible we shall use the term 'self-active' phantasy, calling Jung's 'active phantasy' (*q.v.*), 'imaginative'.

activity (in its widest sense),[1] where they are called forth intuitively, consciously selected and finally integrated with the relevant cognitive contents. The aim of this paper is to endeavour to demonstrate this process and to trace the steps by which it occurs. To this end, a study was made of the cognition and phantasy of children ranging in age from seven to twelve or thirteen—the age range during which a transition from the 'self-active' to the imaginatively used phantasy may be expected to occur. Moreover, a number of children attending psychological clinics were also studied. It is clear from the work of Lowenfeld (7) that the phantasies of these children frequently do not develop in a normal way; and it seemed that this abnormal development, and its consequences, might throw some light on the normal processes.

The cognitive and phantasy processes to be studied centred in the main about the perception of rather complex pictures, and the verbal response given when the child was asked: "Tell me about this picture." This type of activity was chosen because it provided an objective control of the cognitive (perceptual) processes, but at the same time gave scope for imaginative activity. Also, as we shall see below, it often allowed the appearance of the childish 'self-active' phantasy. Certain other types of activity were also studied, to supplement the above. (1) and (3) were introduced to give further opportunities for the expression of phantasy; (2) was intended as a check on the child's cognitive abilities in non-verbal activity:

(1) Verbal description from immediate memory of visually perceived meaningless, non-representational presentations;

(2) Reproduction of pictures from immediate memory, by arranging figures in the positions occupied by them in the original presentation;

(3) Free construction of pictures, i.e. the child arranged figures to make an imaginary picture.

II. METHOD OF EXPERIMENT

There were two main experiments, carried out with different groups of ordinary school children and of children attending psychological clinics for some form of emotional maladjustment or behaviour difficulty.

Experiment A. In the first experiment, a series of coloured designs, each 5×4 in., was presented to each child by means of a tachistoscope. This instrument is fully described in the Appendix to Part II of this paper. Here it suffices to say that it was of the Dodge type (4); and that

[1] It would seem possible that the 'school-teaching', 'home-making' and 'doll-nursing' type of play mentioned above may show one form of this imaginative activity.

the designs were seen binocularly, for a duration of 1·8 sec., at a distance of about 2 ft. Rigid fixation and an exact position of the head were unnecessary, but the child was given a warning signal before each presentation. He was instructed to "tell me about these pictures", and his responses were recorded verbatim.

The series consisted of six groups, lettered A–F, each of four designs numbered 1–4. In each group, 1 was a meaningful representation of a scene of real objects; 2 was a semi-diagrammatic representation of the same scene in approximately the same colours; 3 was much more distorted and meaningless, and usually differently coloured, but still bore some resemblance to 1; 4 consisted of irregularly shaped patches of colour. Fig. 1 is an uncoloured reproduction of B1–4. The other meaningful pictures were as follows:

A: two boys running along a path down a hill;

C: a house with a wall and a hedge round it and a tree on either side of it;

D: a harbour with a jetty and lighthouse, and four sailing boats on the sea;

E: a man punting a boat across a river;

F: two archers in green mediaeval costume shooting.

To half the children (approximately) the designs were presented in the order: A1, 2, 3, 4, B4, 3, 2, 1, C1, 2, 3, 4, D4, 3, 2, 1, E1, 2, 3, 4, F4, 3, 2, 1; and to the other half in the reverse order. Thus the responses to half of the semi-meaningful and meaningless designs were to some extent suggested by the preceding meaningful designs; while with the other half this suggestion was lacking.

The subjects were: 20 children aged 7–8 (group I) and 20 aged 10–11 (group II) from a secondary school; and 12 aged 7–9 (group I) and 15 aged 10–13 (group II) of children attending a psychological clinic. The former carried out the experiments at their school, the latter at the clinic.

Experiment B. In the main part of the experiment, there was the following procedure:

(1) A large coloured picture, 25 by 20 in., of a scene in a farmyard (called the *Farmyard*) was presented to the child, who was instructed: "I want you to look at this picture and tell me all you can about it." All responses to this and subsequent presentations were recorded verbatim.

(2) The child was instructed: "I am going to show you a picture for a short time, and I want you to look at it carefully and tell me *afterwards* all you can remember about it." He was then shown a picture of similar size of a quarrel in a public house (called the *Tavern Brawl*), for a period of $\frac{1}{2}$ min.

(3) The child was instructed: "I am going to show you another picture for a short time, and afterwards I am going to give you some little figures just like the ones in the picture; and I want you to arrange them on a piece of paper in the same places as they were in, in the picture." He was shown a picture of a milkman's horse running away (called the *Runaway Horse*) for a period of $\frac{1}{2}$ min. Then he was given painted cardboard figures, replicas of those in the picture, and a sheet of paper painted in the same way as the background of the picture. After he had arranged the figures, he was instructed: "Now tell me about this picture."

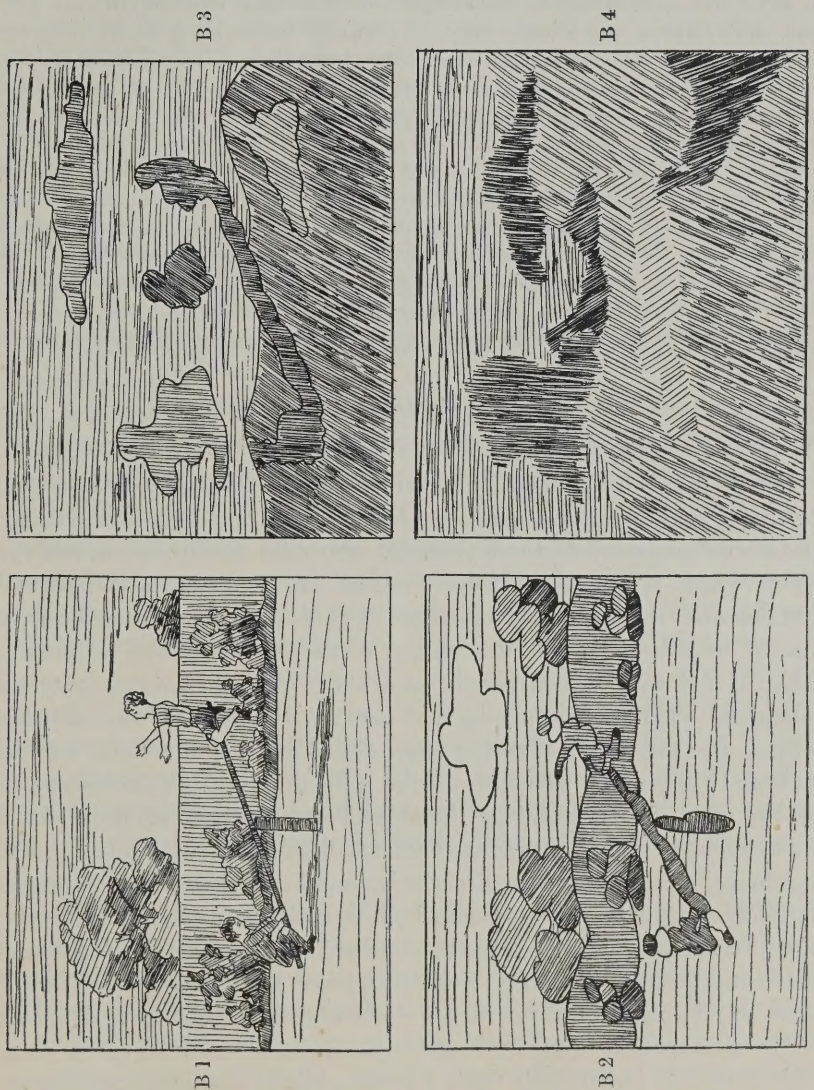


Fig. 1.

(4) The child was instructed: "Now I am going to give you some more little figures and another piece of paper, and I want you to arrange them so as to make a picture—any kind of picture you like. You needn't use all the figures—just take the ones you want." He was then given a sheet of paper painted to resemble a green field with a blue sky above, and painted cardboard figures of: Five white men (two of them with rifles), three Indians, two women, three children, four horses, six dogs, six farmyard animals, seven wild animals and six trees. (Some impression of these animals and the way in which they were arranged can be gained from Fig. 2, p. 291.) After the child had arranged them to his satisfaction (in what is referred to subsequently as a *free construction*), he was instructed: "Now tell me about it; what is it a picture of?"

The subjects were: 28 children aged 11 from a county senior school, and 14 children aged 9–11 (also one of 7 and another of 13) from a psychological clinic.

In a preliminary experiment, a picture of a medieval stag-hunt was used instead of the Runaway Horse for memory reconstruction (3); and the Runaway Horse was used for description from memory (2). The subjects were five 7-year-olds, ten 8-year-olds and six 9-year-olds from a county junior and infants' school.

These experiments were conducted in the children's schools or clinic respectively.

III. Normal cognition and phantasy

(1) *Perception and immediate recall of meaningful pictures*

It was found by Binet(2) that three successive stages could be distinguished in the verbal responses of the child to the perception of fairly complicated pictures. At three years old, the child merely *enumerated* the objects which he observed in the picture. At seven years the normal child *described* the picture and its more obvious qualities; and as he grew older he included a simple description of the actions of the people depicted. But it was not until fifteen years of age that the child could really understand what was happening and *interpret* the picture fully. Binet's pictures, however, were of rather unusual and distressing scenes, the full emotional content of which would be difficult for a pre-adolescent to understand and interpret. In the latest Terman-Merrill revision of the Binet-Simon tests(9), pictures are used which are more within the child's experience, and enumeration, description and interpretation are expected from the normal child of three-and-a-half, six and twelve years respectively. Ségers(8), using a picture of a sick child, its mother and the doctor, found that interpretation was fairly well established at eleven years. Table I shows Séger's figures for the number of children at various ages giving simple enumeration, complex enumeration (characterized by more complete responses occasionally related to each other), description and interpretation. Between six and ten years, the descriptions became generally fuller; figures of people were more clearly distinguished, and the background and irrelevant details, often described by the younger children,

Table I

Age in years	No. of children giving: Simple enumerations	Complex enumerations	Description	Interpretation
3– 4	10	0	0	0
4– 5	9	1	0	0
5– 6	4	6	0	0
6– 7	2	2	6	0
7– 8	3	2	5	0
8– 9	2	1	7	0
9–10	1	0	7	2
10–11	1	2	6	1
11–12	0	2	1	7

were omitted. Though the number of children studied was small, it is clear from the Table that Ségers's figures agree well with the Terman-Merrill revision. There will be little interpretation before ten years, and an appreciable number of children will continue to enumerate until eight or nine, and even older.

It might therefore be expected that *enumeration* would occur among the younger children described in this paper, though it would not be common; and that interpretation would begin to appear among the older ones. The first expectation is fulfilled by the school children[1] of Exp. A. In three cases there was a fair amount of enumeration, but it was fairly complex: "A house with a wall and a gate and a path"; and description also occurred: "Two little boys running down a hill." Among the school children in Exp. B enumeration was more common; 50% even of the eleven-year-olds gave enumerative responses. This result was partly due to the fact that the intelligence of the Exp. B school children was lower on the average than that of the Exp. A school children. But ten of the fourteen enumerators in Exp. B had been marked A on an intelligence test; whereas of the nine children in the group who had been marked B, only four were enumerators. Clearly, then, other factors were at work besides intelligence.

It appears from Table II that the response was to a considerable extent a function of the picture shown. Picture (1), the Farmyard, was a rather static picture, lacking in exciting incident. Picture (2), the Tavern Brawl, for the eleven-year-olds, and the Runaway Horse for the younger children, was not only more exciting and dramatic; but also it depicted a central unifying incident, rather than a collection of minor details. The children were interested in the picture, and their interest aroused a more

[1] The clinic children also attended school; but it is convenient to refer to the non-clinic children as 'the school children'.

intelligent manner of response. Moreover, they were required to respond to picture (2) from memory and picture (1) from immediate observation. Thus even the slower and lazier child was obliged to formulate his response in the manner most easily remembered; and probably a brief description clinched his recollections more clearly in his mind than did a list of minor details.

Table II

Age in years	No. in group	No. of enumerative responses to Picture (1) No.	Picture (1) %	Picture (2) No.	Picture (2) %
7	5	4	80	2	40
8	10	8	80	3	30
9	6	3	50	0	0
11	28	14	50	3	11

We may conclude, then, that the majority of children of these ages (see Table II) will give to a normally interesting picture, recalled immediately, a descriptive rather than an enumerative response; though there is a tendency for the slower and lazier child to resort to enumeration when neither his interest nor his powers of observation and recall are aroused. The eight children who gave enumerative responses to picture (2) as well as to picture (1) were either of low intelligence, or showed definite signs of emotional maladjustment such as will be described in § IV in Part II. For the moment it suffices to say that enumerative responses in children of these ages may be a function of emotional maladjustment, rather than of low intelligence.

It has already been noted that as the child grows older his *descriptions* become fuller, more accurate, with more description of movement, and the irrelevant detail becomes suppressed. There is, however, one particular type of description which should be noted, because it is frequent even among normal children. This for the sake of convenience is called *designation*, though perhaps it approaches the *syncretism* described by Claparède(3). I have called it designation because it gives the impression that the child is simply giving a title to the picture. It may consist of a single word, e.g. "garden", "snakes", "hill"; or of a short phrase: "Some little boys running". "Some black-coloured hills". It gives the hearer the impression that the child has, so to speak, put the matter in a nutshell; the subject is exhausted and there is no more to say about it. Whereas in complex enumeration, which it resembles to some extent, the child will drift on from one detail to another; there seems no reason why he should stop, because he may quite well begin again at the beginning and carefully repeat himself.

In Exp. A eight of the younger school children (group I) and eight of the older (group II) (40 % of each group) gave a considerable number of designatory responses. This type of response was not merely a function of intelligence, since some of these children were of good intelligence, and others rather below normal. It can hardly be called an abnormal type of response, since it occurred with such frequency. At the same time, it appeared possible that it was connected with a definite inhibition of cognitive activity, caused in some cases by mental or physical subnormality, but in others by emotional resistances of an aggressive kind. We shall return to the subject in § IV.

It is not easy to establish the point at which description is superseded by *interpretation*. It was found almost impossible to draw a hard and fast line with the rather simple meaningful pictures of Exp. A. The activities of the people in these pictures were clear and obvious, and a description of them generally included such interpretation as was possible. Certain of the clinic children, it is true, ascribed emotions and emotional expression to some of the people in the pictures, but these seemed mostly to be pure inventions—there was nothing in the pictures themselves to suggest them. However there was one type of response, that given by Katharine H. (see p. 287), which seemed to suggest the beginning of interpretation; we shall return to that shortly.

More elaborate pictures were shown in Exp. B. It is true that the picture of the Farmyard did not perhaps give much scope for interpretation, though it was expected that an intelligent child might point out that it represented an evening scene, with the cows coming home to be milked and the horses to be stabled; a scene very familiar to the country school children. In point of fact, only one response seemed to be of this kind: "Looks like a farmyard picture; . . . the man looks as if he's going to feed the cows; one of the boys is going to take the cows in; the horses look as if they're coming home from work" (May C., age 11). Five other eleven-year-olds gave responses which showed attempts to expand and interpret part of the picture, e.g. "the man's going to milk the cows", "men taking the horses back to the stables", etc. Of the younger children, three nine-year-olds and two eight-year-olds also gave partial interpretations.

The kind of interpretation that might be expected from an adult of the Tavern Brawl would be:

> It's a public house; two men (dagoes) have started a drunken quarrel and are going to attack each other with knives, but they are forcibly restrained by three sailors; the barman and barmaid are much alarmed by what is happening.

None of the children gave as coherent an interpretation as this; the best was given by one of the clinic children, Joan M. (age 11–6, I.Q. 105):

It's a public house; lady was serving behind the bar; two gipsies with knives causing trouble; three more men settling the trouble and going to throw them out; they're throwing glasses, and the beer's upset[1] on the floor, and the stool's upset as well; a bottle on the counter; men that are helping to settle the trouble got high boots on; lady behind the counter got a shawl on and looks worried.

Thirteen of the school children (46 %) gave partial interpretations which were less complete and coherent than this. Thus it appeared that these children, even though they had some understanding of the activities represented in this picture and could give partial interpretations of them, were not able to group their impressions in such a way as to obtain a general understanding and give an interpretation of the picture as a whole.

The expected interpretation of the Runaway Horse would be:

The milkman has been delivering milk and leaving his horse and cart standing in the road; the horse has been frightened by a barking dog and has bolted, with the milkman running after him and trying to catch him; two children playing in the road have just managed to get clear, although the little girl has tumbled in the gutter; just at this moment a youth has rushed out of the house and is stopping the horse.

Here five of the clinic children had a fairly general understanding of the picture. The best interpretation was given by Stanley H. (age 10–1, I.Q. 96):

A milk cart and horse's been frightened away down the street and the milkman's running after him; a lady is looking out of the window, and a lady is coming out of the door, and a man come out of the door; he's stopped the horse; little girl's hurt herself; the dog kept on barking and made the horse run; the other boy's running away in case he might get hurt by the horse.

The older school children did not describe this picture. But the younger ones seemed to understand fairly well that the milkman's horse was running away, though rather fewer mentioned the fact that the man was stopping it. It appears, then, that with a suitable type of picture interpretation of a kind can begin at ten or eleven.

But one is struck by the fact that even children such as Stanley H., who seemed to understand the picture to the extent of knowing what was happening in it, did not, as most adults would, begin with the essential points and add the less important ones later. Thus with the Tavern Brawl, spilt beer and overturned tables and chairs were more emphasized than the men fighting. Again with the Runaway Horse, the women

[1] The 'spilt beer' was mentioned by most of the children, and seemed to impress them more than anything else in the picture.

looking out of the houses were frequently mentioned before the man stopping the horse—and indeed he was sometimes omitted altogether. It seems that the child has some general idea of what is going on, but becomes confused by the detail and cannot formulate the essential points clearly; so that his response can hardly be called 'interpretation' in Binet's sense of the word. Certainly none of these responses were as advanced and coherent as that quoted by Ségers ((8), p. 729) for an eleven-year-old child. On the other hand the children's impressions of the picture were usually clear and accurate enough for them to reconstruct the picture from memory. Only five of the eleven-year-old school children (18 %) placed the figures in such a way in their reconstructions as to show they quite obviously had not grasped the essential features of the picture. It seems probable that the child when responding verbally is confused between what he actually perceives and by the phantasied elaborations of the cognized theme, as to the further activities and the emotions of the people depicted. He has not yet reached the stage at which he can make appropriate selections from the contents of his phantasies, and combine them suitably with cognitive material to produce a coherent 'interpretation' of the picture as a whole.

(2) *Perception of meaningless designs*

It has been shown that the majority of the school children in both age groups used some type of description in responding to the meaningful pictures in Exp. A. The descriptions were much fuller in some cases than in others, containing sometimes more description of activities, in others more details about colour, objects in the background, etc. But these descriptions were accurate in that no actual inventions were introduced by these children. We shall see in Part II of this paper that it was characteristic of some of the clinic children to introduce inventions[1] in responding to meaningful pictures.

But when the designs were more or less meaningless and did not represent real objects and people, a great deal of invention occurred, in fact clearly it was the normal type of response to this kind of material. Particularly was this the case with the designs numbered 2 which were only semi-diagrammatic, and did resemble pictures of real objects. It will be seen from Table III that almost all the children named some object

[1] The word 'invention' is used for any feature mentioned in the response which was not derived entirely from the immediate objective presentation. Thus in certain cases inventions were wholly or in part derived from recollections of previous percepts, while in others they were completely spontaneous and original (see below).

or person suggested to them by presentations of this kind. And even with the much more diagrammatic designs of the kind numbered 3 and 4, less than 20% of the responses were purely objective descriptions of shapes and colours, with no allusion to anything representational. The older children in group II were slightly, but only very slightly, more accurate in this respect. However, it did appear that the "effort after meaning" (Bartlett(1)) decreased somewhat with the less pictorial designs.

Table III

	% of responses containing inventions to designs numbered		
Group	2	3	4
I	98	84	81
II	98	82	78

As with the meaningful pictures, we find the younger children responding enumeratively with strings of names: "trees, grass, rocks, clouds", etc. The designs numbered 2 and 3 were often called "men", "boys", "people", "figures", "animals". As the children grew older the responses became more descriptive and simple actions were assigned to the people and animals. These designs did of course resemble the meaningful pictures in the series, so that to some extent people and actions were suggested beforehand. But it did seem that much spontaneous invention was occurring which was not merely, or mainly, suggested by the meaningful pictures, but proceeded from subjective phantasies. It was possible to differentiate roughly between those inventions which were 'derived' directly from the meaningful pictures which immediately preceded them; those which were 'similar' to the objects in the meaningful picture of the same group, although the child did not see the latter until afterwards; and those which were entirely 'original'. The numbers are given in Table IV. It is clear that with the presentations numbered 2, 'derived' and 'similar' responses were much more frequent than 'original'; with the more diagrammatic presentations, 3 and 4, 'original' interpretations became more frequent than the other types of invention. But we cannot be certain that even the 'derived' inventions were entirely free from 'original' invention and based purely upon the child's recollections of previous percepts; indeed it is much more probable that spontaneous elaboration did enter into them to some extent. Particularly the tendency to invent men and animals seemed to operate almost throughout. Table V shows the number of responses containing invented men, animals and animated objects such as "ghosts", "scarecrows", moving

Table IV

Design no.	No. of inventions per child in Group I: Original	Similar	Derived	Total	Group II: Original	Similar	Derived	Total
2	1·9	6·6	6·0	14·5	1·4	7·6	6·1	15·1
3	4·4	2·1	3·1	9·6	4·4	2·0	3·1	9·5
4	6·5	0·1	0·5	7·1	7·2	0·2	0·6	8·0
Total for school children	12·8	8·8	9·6	31·2	13·0	9·8	9·8	32·6
Total for clinic children	17·3	9·3	13·4	40·0	12·0	10·3	11·4	33·7

Table V

Design no.	% of responses containing inventions of men, animals, etc., to design lettered: A	B	C	D	E	F
			Group I			
2	95	70	0	0	100	100
3	70	0	0	0	25	65
4	45	0	55	0	15	50
			Group II			
2	100	90	0	0	100	100
3	90	15	0	0	45	85
4	35	0	30	5	35	55

"things" and "figures". The only group to produce no inventions of this kind was the D group. D 1 was a picture of boats in a harbour, D 2 and 3 showed geometrical shapes and patches of various kinds, and D 4 contained irregular splashes of colour which apparently did not suggest any animal shapes—though the response, "a lot of washing blown up high", shows human influence. (The response to D 3, "a flood—just the tops of houses", demonstrates clearly the compromise between the percept of D 3, the recollection of D 1, and phantasy elaboration.) Again, although C 2 and 3 were geometric in design and produced no animal inventions, C 4 contained twirly black lines which suggested "snakes", "monsters", "dragons" and "ghosts". These and other similar inventions apparently proceeded not so much from recollections of things previously perceived and experienced, but from phantasies of rather uncanny animals, moving in rather a frightening manner; such phantasies seemed ever present with many of these children. Thus "snakes", "dragons" and "ghosts" were "having a fight"; there were "two blackies—one's jumping over the other"; "a quarrel, with black men in"; "a snake on the hill and a swan flying, and the snake was jumping up to try and kill it". Even when the child could not perceive anything in the objective presentation which could be definitely named, he gave reports such as: "black things, jump-

ing by some rocks and banks"; "black things jumping over each other ; "somebody had got a cover over them"; "something sitting on a mushroom"; "black horse thing—and two people riding it"; "they're looking at something".

Now it is clear from these responses that the children did not introduce as inventions merely the ordinary men and animals and activities familiar in their everyday life. There was obviously a suggestion of 'magic'—of creatures and activities, more familiar in dreams than in waking life. The child's 'self-active' phantasies seemed to be waiting ready to be introduced in any favourable situation with an appropriate setting. The amount of phantasy production was kept within limits by what he actually perceived. Thus both the number and nature of these phantastic inventions were limited. Very few of them seemed to have lost all trace of resemblance to the original presentation, though Michael W.'s response to B 2 (see Fig. 1) came near it: "An aeroplane and three little men; green and all colours, and a path; they've got a piece of rope and it was tied on the tail of an aeroplane—I think it was an aeroplane, or the shadow of an aeroplane." We shall find more responses of this kind among the clinic children. But since the phantasy material was in general limited, it is not possible at this stage to analyse its nature and content in any detail. It can only be said that:

(*a*) It deals in the main with living creatures. These are sometimes people or familiar animals; but often more or less 'magical' animals (snakes, ghosts, dragons, etc.); and sometimes mysterious and nameless "things" or "figures".

(*b*) These creatures are often engaged in some rather dangerous or suspicious occupation—"fighting", "cutting each other", "jumping up at" or "over each other", "running after each other", or just "looking at each other".

But most of the children realized to some extent that they were introducing inventions into their responses, and that the latter were not exact accounts of the objective presentations. It is true that one or two of the younger and least intelligent children produced straightforward inventions without any verbal qualification to indicate that they were aware of anything incongruous or not strictly objective in their response. Two others also gave straightforward inventions, but these seemed to proceed from superficiality of observation, resulting from lack of effort or of co-operation; that is to say, the perceptual element was reduced to a minimum and the response was based almost entirely upon phantasy. This phenomenon appeared more clearly among the clinic children.

The slightly older children qualified their inventions by using the terms: "Some very funny things"; "queer"; "crinkly"; "squiggly"; etc. This seemed to indicate their recognition that the objective presentation did not contain exact representations of the men and animals which they reported. At a rather older stage "like" and "sort of" were introduced as qualifications, and later the response might take the form of a query: "Was it a pathway up by the hill?" Or sometimes the response indicated that perception and phantasy existed side by side, but separated, with no interrelationship between them; thus: "Another blue sky and some black streaky stuff and lots of other colours all round"; "yellow and red and brown and black and all colours, crinkly things, like dragons again".

There were those who introduced little or no invention of any kind. Two such cases were superficial observers who responded perfunctorily with "design", "pattern", "blocks", "lines", etc. Some of the older and more intelligent children, particularly if they were of a rather suspicious and critical turn of mind, called the designs "bad painting", "just a muddle of colours", "painted picture—hadn't been finished", "child's drawing". This seemed to indicate that there was a tendency to introduce phantasy which was repressed as being incongruous with actual perception; but this tendency was still sufficiently strong to inhibit clear perception of the objective presentation. There were, however, two children who seemed to have repressed almost all phantasy, and gave long and accurate descriptions of the colours and shapes in the presentations. Such children were of the careful, methodical and 'scientific' type; but at the same time they showed tendencies to over-conscientiousness and a rather compulsive use of verbalization. We shall return to this type later.

Finally there was one child who gave evidence of a really satisfactory type of reconciliation between perception and phantasy. Katharine H. (age 12–2, I.Q. over 136) was one of the clinic children, not one of the ordinary school children. She alone realized that the designs formed a progressive series from meaningful pictures through semi-meaningless designs to patches of colour. She understood the way in which successive presentations became more or less meaningful, thus: (A 1–3) "Two boys running down a hill"; "The boys were changed and all squiggled"; "Looks as if they were beginning to turn into trees". One might say that she had reached the stage of real interpretation. For here the objective configuration and the imaginative invention were reconciled and synthesized, on a basis of rational understanding.

(3) *Free construction*

So far we have dealt with cognitive activities which were more or less closely constrained by the necessity of reproducing some particular objective presentation. But in Exp. B the school and clinic children were given an opportunity of exercising both their purely cognitive abilities and also their phantasies in a less limited fashion. It is true that limitations were imposed by the nature of the material given them for construction. But the actual selection and arrangement of this material were determined by their tendencies to invention and by the boundaries imposed by ideas of reality derived from knowledge and understanding of the world outside.

There appeared to be three main stages in the development of these constructions, in general succeeding each other as the child's age increased:

A. The youngest school children seemed to make no real attempt at construction, but to endeavour to crowd as many of the figures as possible into the given space. This was not done simply at random, but quite carefully, apparently with some idea of pattern in the arrangement (i.e. the figures were sometimes placed in long straight rows across the background). But it was not possible to see any ulterior significance in this pattern. Thus A seemed to correspond to the stage of enumeration in perception. Four of the seven-year-olds, two of the eight-year-olds and one nine-year-old (one-third of the younger school children) produced such constructions.

B. At the next stage, selection of some kind took place. Either a large class of figures was rejected, e.g. all the wild animals, or certain members of each class were selected, and these were grouped in a way which was meaningful to the child, and which he could sometimes describe verbally. Sometimes the significance of the grouping was fairly obvious to the experimenter, at other times more difficult to grasp. But the arrangement did not form a coherent whole; it was rather a collection of segregated figures or groups, each group varying in number from two figures upwards. This stage seems to correspond to description and partial interpretation in perception. Eleven (42 %) of the younger school children were at this stage, and nineteen (68 %) of the eleven-year-olds.

C. At the highest stage the figures were arranged in a single coherent whole—what in fact we should call 'a picture'; and this was given a name and a rational description. Three (14 %) of the younger school children and nine (32 %) of the eleven-year-olds had reached this stage.

It seemed that the transition from stage A to stage B was mainly a function of mental age. At stage A the child's phantasies were too incoherent or insufficiently related to his cognitive functions to find expression in this form of free construction. Either the child was not fully aware of his phantasies, or his phantasies were dissociated from his cognitive activities and the latter insufficiently developed for utilization in dealing with the material provided.

At stage B the child was able to make use of his knowledge of the real world in cognizing the material, and his cognitive ability in manipulating it; also he could use the latter to express his phantasies. But he could not yet make use of phantasy material actively, in unified imaginative construction. Thus we find that some children inclined more to the use of isolated pieces of recollected experience of the familiar external world. Others used the experimental material more or less in symbolic form, to express their phantasies; while yet others seemed to hover uneasily between the two modes of expression. From the first type came productions which appeared similar, and were described in similar fashion, to the partial interpretations which were made by the child in his observation of and response to pictures. These consisted of a number of groups of people, animals, etc., each of which had a meaningful relation. Moreover, there was a vague coherence or 'fittingness' between the groups—that is to say, they were not wildly inconsistent with each other; but there was no complete and logical relationship between them. They were described in a set of consecutive but unrelated verbal statements; thus Arthur H. among the eleven-year-olds gave:

> In the field all the cattle feeding and racing about. A farmer walking with a little girl and a dog. A man shooting with another dog. Two horses racing along. Two pigs fighting each other. Two sheep and a man walking.

Here it appears that the child had a general idea of a country-side with people and farm animals engaged in normal activities; but these were only very loosely interrelated. No imaginative process had united the separate items into an interpretative whole.

But there were frequent cases in which there was considerable inconsistency between the separate groups. Thus Douglas P. gave:

> Driving some horses down to a field. A dog chasing a boy. A boy with his elephant. A man and a woman talking. A boy calling his dogs.

This type of description was sometimes greatly elaborated with much

more detail and a larger number of segregated groups. Sometimes it was more obviously incoherent, as in the case of Rosemary P.:

A man running after a horse trying to catch it; some people watching on. A man trying to shoot something[1] near the trees. A boy running away from it. A lady going to market.

Again in some cases the child found it easier to give a general name such as 'Zoo' or 'Jungle', to a number of unrelated groups. Thus John P. produced seven or eight groups, of men shooting animals, dogs chasing sheep and cows, a man chasing a horse, a child with a woman and a dog, etc.; and this production was called, after much hesitation, "Game hunting". It seemed probable that some of these items were suggested from previous experiences, and were based upon the perception and cognizing of something existing in the external world. Other items had arisen in phantasy, possibly from association to the more normal items. But logical coherence and interrelationship were lacking because the child was unable to make active use of the phantasy elements.

There were other cases in which there seemed to be a definite line, or lines, of demarcation between different parts of the construction. Thus Kenneth B. described his construction as:

Wild animals at the back, and a man going after horses, and some sheep and cows and people going across the field.

The wild animals, and four men attacking them, were arranged as a horizontal strip across the back of the field; in front of this in the centre was the man going after the horses; on one side, a man, woman and child and two dogs; on the other a man, a dog and some sheep. Here it appeared that the phantasies about dangerous wild animals and the attack on them were pushed to the back of the mind, while ordinary activities (man and horses) and peaceful occupations (people, dogs, sheep) were segregated from them in the foreground. In other cases the isolation of the phantasied danger was evident from the fact that the child described the rest of the construction, but made no mention of what was clearly an important part of it. Thus Leonard B. called his production, "Meadows with sheep and horses and a man about", but did not allude to a man shooting a bear, in the middle of the back of the field. Again Bessie M. described hers as, "People have come to see that girl; and that's the keeper with his dog", making no mention of the line of wild animals placed across the field behind the people and the dog.

[1] A rather large and ferocious looking cat, several times taken for a wild animal.

Finally we may include among the stage B productions one by Monica B. in which the dangerous phantasies seemed to be gaining the upper hand. It was called simply, "Jungle, with some animals in it"; but the wild animals were arranged in a semi-circle round the lonely figure of a little girl of about the same age as Monica!

This construction might almost be grouped among the unified constructions of stage C, but that it showed a clear lack of contact with reality. All those assigned to stage C were to some extent 'real'-ized,

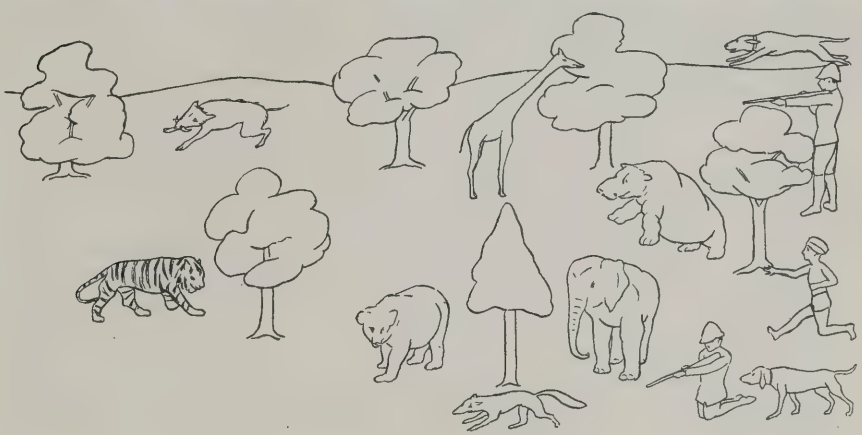


Fig. 2.

although they may be divided into those which were mainly derived from objective experience, and those which contained a large imaginative component. Of the former, the most intelligent and best arranged was made by Kathleen B., who described it as, "Field with some lambs and some horses, and a man going to bring in the horses". But in the main these productions were rather pedestrian and unimaginative, and possibly derived to some extent from the picture of the Farmyard, which the children had seen previously. Indeed it is possible that some of them contained so little of the unifying imaginative factor that they should have been excluded from this stage. The rather popular 'Jungle' groups gave more scope for imagination. These in general took the form of a group of men shooting or chasing wild animals. But sometimes the men were right in one corner of the picture (Tony P. (reproduced roughly in Fig. 2), John C.); and sometimes they were themselves being chased from behind by other animals (Stanley N.). Here clearly the dangerous and fear-provoking forces were in the ascendent. Sometimes the forces were about equally balanced, however (Derek W., Vincent B.). It may be

noted that Stanley N. was the only one of these children to mention that the men were themselves being attacked by animals. There were also two cases in which the child's own aggressiveness seemed to be finding expression. Thus Donald M. made a "field with horses, a man waving a whip and frightening them, with a dog"; and Rosemary J. produced: "A fox, and some men chasing it" (there were five men, a child and three dogs against one fox).

It is possible to compare stage C in free construction with the interpretative stage of perception, though at first there may appear to be little justification for this comparison, since in interpretation the child is compelled to consider carefully the nature of the objective presentation in order to interpret it correctly; while in free construction he can use his own imagination freely. But we have already pointed out that even in free construction the child may definitely be limited by his ideas of reality, i.e. of the kind of activities and situations which are characteristic of people and wild and tame animals in the real world. Moreover the child at stage C is not only observing these limitations, but has also synthesized his construction into a coherent whole. In the same way when he gives an interpretative response to a perceived picture he has 'imagined' into it ideas, situations, activities, which are not actually represented there; he has selected the appropriate ones in the light of his experience of the real world, and has synthesized them into a coherent whole. Thus we conclude that the characteristics of these stages which demonstrate the functioning of the fully developed powers of imaginative interpretation are:

(1) The accurate perception and cognition of the objective situation, i.e. what is normally called perception (see Vernon(10), pp. 5, 6);

(2) The phantasying of more or less emotionally charged situations suggested by, and associated from, the percept;

(3) The active selection from this phantasy material of what is relevant to the percept *and* what is consistent with knowledge and experience of the real world (i.e. in accordance with the 'reality principle');

(4) The final synthesis of cognized and phantasy material so that the response consists of an integrated whole containing what is essential and relevant both of the immediate percept and of the appropriate phantasy material; less important features being relegated to the background and unimportant detail being altogether omitted.

In perception of a picture the response is determined to a considerable extent by (1). In free construction, (1) is limited to the perception of the

figures and the background sheet of paper, and the response is determined to a much greater extent by (2). But in both cases, (3) and (4) are essential in a response at the stage C or interpretative level.

Stage B and the level of description and partial interpretation are generally characterized by (1) and (2) above, but (3) and (4) are more or less incomplete. That is to say, as a rule perception is fairly accurate, and can be reported clearly, as in description. 'Self-active' phantasy also occurs when there are inventions, and in partial interpretation; this can be described verbally, but cannot be unified with the perceptual material. The response may thus contain perceived and phantasied material side by side, either unrelated or in conflict with each other. Since the phantasy is 'self-active', it is not selected or controlled cognitively; thus no real synthesis between cognitive and phantasy elements can occur.

But it appeared that the absence of imaginative interpretation was to some extent a function of the objective presentation. Clearly it would be difficult for even an intelligent adult to produce an imaginative interpretation of a picture which was in the main a collection of minor details, and gave little opportunity for the exercise of imagination. And with such a picture, the child may confine himself to objective description. Or such description may occur when the child is for some reason or other unable to express his phantasies, or has completely repressed them. On the other hand there are cases, and we shall describe more of them in Section IV of Part II, where accurate perception and contact with reality are at a minimum, and the response is mainly the result of 'self-active' phantasies, containing little of the perceptual element. Thus the appearance of descriptive responses or stage B constructions may be due to:

(*a*) Lack of maturation through age; i.e. they are characteristic of the normal child of less than about ten years;

(*b*) Unsuitable perceptual material;

(*c*) Failure of cognitive ability or phantasy, or both, to develop normally through some emotional maladjustment.

It is less easy to understand the nature of the response at the stage A and enumerative levels. Verbal communication is difficult at this level. Thus we cannot be sure if the phantasies are themselves so incoherent that the child is not fully aware of them; or if the cognitive abilities are insufficiently developed to be used in expressing them. Clearly the careful but apparently meaningless arrangement of objects in free construction is closely analogous to the listing of names of objects in enumeration. We can only conclude that the child is trying out its recently developed

cognitive skills (since naming objects is the earliest form of language) without at present being able to use them for the expression of symbolically imaged phantasies. As soon as the child begins to draw inferences from its observations as to the actions and emotions of the people, etc., depicted in a static picture, and to formulate these in words, a real link has been formed between conscious cognitive activities and phantasy processes.

It may be objected that the five-year-old children studied by Griffiths (5) were able to express their phantasies in words and in other activities such as drawing. Thus it seems that the child *can* use such techniques for expressing phantasy even when quite little. But apparently only an older child can use them to describe synthesized phantasy and cognitive material. The phantasy material can appear in words and in constructive activities, but it is much more difficult for the child to use it deliberately, to employ it consciously and only when it is fully congruent with his perceptions and knowledge of reality. We may suppose, and this would agree with the conclusions of Griffiths, that it is by expressing his 'self-active' phantasies in words (such as we have encountered in the 'partial' interpretations) that the child becomes more fully conscious of them; and is thus enabled to observe them, and finally to select and make use of them in imaginative interpretation and construction.

REFERENCES

(1) Bartlett, F. C. (1916). *Brit. J. Psychol.* viii, 222.
(2) Binet, A. & Simon, Th. (1908). *Ann. Psychol.* xiv, 1.
(3) Claparède, E. (1916). *Psychologie de l'enfant et pédagogie expérimentale.* Geneva.
(4) Dodge, R. (1907). *Psychol. Bull.* iv, 10.
(5) Griffiths, R. (1935). *Imagination in Early Childhood.* London: Kegan Paul.
(6) Jung, C. G. (1923). *Psychological Types.* London: Kegan Paul.
(7) Lowenfeld, M. (1935). *Play in Childhood.* London: Gollancz.
(8) Ségers, J. E. (1926). *J. de Psychol.* xxiii, 723.
(9) Terman, L. M. & Merrill, M. A. (1937). *Measuring Intelligence.* London: Harrap.
(10) Vernon, M. D. (1937). *Visual Perception.* Camb. Univ. Press.

SOME EXAMPLES OF CONSTRUCTIVE THINKING AMONGST GREENLANDERS

By D. M. CARMICHAEL

From the Cambridge Psychological Laboratory

I. *Place and type of subject* (pp. 295–296).
II. *Aim of the experiments* (pp. 296–297).
III. *Material and method* (p. 297).
IV. *Example I* (pp. 297–300).
(i) *Material* (pp. 297–298).
(ii) *Method of presentation* (p. 298).
(iii) *Analysis of results* (pp. 298–300).
(iv) *General conclusion from Example I* (p. 300).
V. *Example II* (pp. 300–304).
(i) *Material* (pp. 300–301).
(ii) *Method of presentation* (p. 301).
(iii) *Analysis of results* (pp. 302–303).
(iv) *General conclusions from Example II* (pp. 303–304).
VI. *Example III* (pp. 304–309).
(i) *Material* (pp. 304–306).
(ii) *Method of presentation* (p. 306).
(iii) *Analysis of results* (pp. 306–308).
(iv) *General conclusions from Example III* (pp. 308–309).
VII. *Example IV* (pp. 309–312).
(i) *Material* (pp. 309–310).
(ii) *Method of presentation* (p. 310).
(iii) *Analysis of results* (pp. 310–312).
(iv) *General conclusions from Example IV* (p. 312).
VIII. *Results* (pp. 312–315).
(i) *The nature of constructive thinking* (pp. 312–313).
(ii) *The method* (pp. 313–315).
References (p. 315).

I. Place and type of subject

The experiments about to be described were carried out among the native population of the Umanak district of West Greenland, during the summer of 1938. In this district, the admixture of European (especially Danish) blood is so strong that it is doubtful whether any pure Eskimos

remain. The natives call themselves "Greenlanders", a term which will be used throughout this paper.

Under the Danish administration, whose ideal is a well-planned and carefully sheltered transition from the original type of Eskimo community to a form of European civilization suited to the country and its inhabitants, all West Greenlanders are Christianized, have churches, schools and a medical service, and enjoy some degree of representative self-government. Most of the subjects who volunteered for these experiments had attended, or were attending, primary schools where reading and writing in the Greenlandic dialect of Eskimo, a little Danish, arithmetic, history and geography are taught, and some, such as administrative employees and 'catechists', or native teachers, had received a secondary education.

The experiments were carried out in Umanak, the administrative centre and chief trading station of the district, and in the two outposts of Igdlorssuit and Qaersut. The two hundred Greenlanders in Umanak are much in contact with white people, for from fifteen to twenty Danes reside there permanently, and in the summer ships from Denmark call at frequent intervals. In each of the outposts, on the other hand, there is only one Dane, the trader, and the majority of the population seldom have dealings with, or even see, any white man other than him.

II. Aim of the experiments

The experiments have two aims, of which the first is the more important:

(i) It is contended by Lévy-Bruhl(1) that primitive people think differently from normal civilized persons, in the sense that their mental processes, as such, are essentially different. The way in which primitive man interprets the universe of his experience is contrasted sharply with the manner in which modern scientists and philosophers deal with this problem, and the conclusion is reached that primitive man's thinking is pre-logical, emotional, mystical and self-contradictory. The very most which Lévy-Bruhl succeeds in proving, however, is that primitive man does not employ certain *logical criteria* of truth and falsity, and no real evidence is brought forward to show that his thinking, as a mental process, is *psychologically* different from that of the scientist or philosopher. The present experiments are designed to investigate whether or not, when normal civilized persons and a relatively primitive people are faced with a certain type of problem, they think in exactly the same way. Lévy-Bruhl, of course, carefully confines his statements to primitives

untouched by European culture. Of the Greenlanders, however, Schultz-Lorenzen has written ((2), p. 210): "The Eskimo mind has kept its peculiar primitive character unchanged...it holds its ideas and groups of ideas as juxtaposed factors. It is unable to join them in a unity.... The Greenlander has no difficulty in keeping in his mind considerable contradiction and contrasts."

(ii) Experiments of a similar type (3,4), carried out with various social groups in England, have shown some interesting differences in the nature, frequency and distribution of conventionalized opinion concerning social behaviour. In some of the Greenland experiments, therefore, the opportunity was taken to invent and use material likely to elicit from the Greenlanders some expression of opinion concerning their attitude to, and working relationships with, Europeans.

III. MATERIAL AND METHOD

The material used consists of a number of "incomplete situations" (see (3)), in each of which the issue turns on the actions of two or more persons or groups of persons. The situations, examples of which are printed below, are left incomplete, but a number of clues are given which indicate possible completions. Subjects are asked to state what they think is most likely to happen in each situation, and their answers are subsequently analysed and classified.

The advantages of this method are: (1) it is *objective*, and does not demand introspection on the part of the subject—an essential requirement when dealing with primitive people; (2) it is *realistic*, presenting situations which demand *practical* rather than abstract scientific thinking, which readily interest all types of subject, and about which most subjects feel justified in expressing an opinion: (3) it eliminates the danger of becoming involved in irrelevant *logical* questions, for it uses situations of which no one of a number of possible solutions is 'more right' than any other. Thus any two solutions may be widely different, or even diametrically opposed to one another, and yet equally relevant to the psychological study of the mental processes from which they result.

IV. EXAMPLE I

(i) *Material*

The Three Hunters

It was winter, and three men were sitting talking in a *fangsthus*[1] far from their settlement. One said: "Joab is a very bad man." "Yes", said another, "since he

[1] Hunting-house.

came to our settlement from South Greenland six months ago he has done all manner of evil things. He has told many lies, he has stolen, he has been cruel to our children, and a nuisance to our women. And only last week he tried to kill my best friend. It is time for us to get rid of him." "That will not be easy", said the third. "Already we have asked the Governor to take him away, but nothing has been done."

For a while they sat silent, and then the man who had first spoken said: "We could kill him." "It is wrong to kill", said the second: "we should be punished." "It would be good for all of us if he were dead", replied the first man: "I myself shall kill him, if ever I can do so without being found out." "We could kill him here, and no one know of it, for this place is very lonely", said the third man. "Joab never comes here", said the second man: but even as he spoke the door opened, and Joab himself entered. "I have lost my sledge and my dogs", he said. The three hunters looked at him, and saw that he was already nearly dead of cold.

The question in this instance is *What will the three hunters do?*

(ii) *Method of presentation*

This story was presented to 100 subjects—43 in Umanak, 36 in Igdlorssuit, and 21 in Qaersut. Of the total number, 53 were men, 21 women and 26 children under the age of 14. In Igdlorssuit each subject was taken individually, the interpreter both reading the story and recording his subject's answer. In Umanak and Qaersut all results were obtained from group experiments, in which the interpreter read the story (usually about two or three times) and each subject wrote down his (or her) own answer. All subjects were allowed to refer to the text of the story if they wished to do so.

(iii) *Analysis of results*

Table I shows the distribution of opinion in each settlement, and for the total number of subjects, in answer to the question, "What will the three hunters do?"

Table I

	Igdlorssuit		Qaersut		Umanak		Total
	No.	%	No.	%	No.	%	no.
Kill	7	19	3	14	12	28	22
Not kill	27	75	17	81	29	67	73
Undecided	2	6	1	5	2	5	5
Total	36		21		43		100

The distribution in each settlement group is remarkably similar. Of the total number of women and children, however, 33% predict that Joab will be killed; of the total number of men, only 10%—which probably reflects a real difference of opinion.

Of those who predicted that Joab would not be killed, many stated positively what the three hunters would do. Their predictions are classi-

fied in Table II without differentiating for sex, age or settlement grouping.

Table II

Help unreservedly	15
Supply with bare necessities and then let him depart	5
Be reconciled	4
Admonish, warn, threaten	4
Punish, beat	2
Consult magistrate	1

One subject distinguished between what *would* happen and what *ought* to happen. "The three hunters", he said, "will agree well enough with Joab, though it is not right that they should do so." Only one subject predicted *individual* rather than *group* action. He thought that the three hunters would feed and clothe Joab, and send him on his way. Then the first man would pursue and slay him, in accordance with his avowed intention.

In addition to stating what would happen, many subjects stated the motives which they thought would determine the line of action taken by the three hunters. These motives are classified in Table III.

Table III

	Igdlorssuit	Qaersut	Umanak	Total
Ethical:				
It is wrong to kill The bad custom of killing has disappeared in Greenland Christian people don't kill	16	2	4	22
It is good to help Greenlanders always help people in distress To help is to act like men Hunters always help each other in times of danger	9	3	—	12
Psychological:				
Joab's plight evokes pity Anger and hatred vanish in time of danger	13	4	7	24

The greater freedom of expression enjoyed by the Igdlorssuit group is clearly evident in Table III and suggests at once that, whenever possible, this method of individual presentation and oral response should be used with relatively primitive people.

What is most remarkable about the Igdlorssuit answers, however, is a 'group self-consciousness' which in some instances amounts almost to 'group self-righteousness'. This self-consciousness appears to a lesser extent in the Qaersut answers, and is entirely absent from the Umanak group.[1] Igdlorssuit subjects, men, women and sometimes even children,

[1] It appears in Umanak, however, in response to specific questioning (see Example III below).

do not say simply, "It is wrong to kill", "Hunters help each other", and so on, but far oftener, "We *know* that it is wrong to kill", "We Greenlanders are Christians and Christians don't kill", "We Greenland hunters help each other", and so on. Indeed, the attribution of irregular behaviour and murderous intentions to Greenlanders brought out from some subjects a real rebuke. One Qaersut hunter says, "Men cannot kill each other—as *you* know also". The head man of Igdlorssuit says, "The beginning of the story is all right, but the rest is bad and thoroughly wrong. I have never heard anything like this in any place". Another old Igdlorssuit hunter welcomes the chance to correct wrong opinion: "If strangers believe such things of Greenlanders, then it is very bad and nonsense. But it is good that I have read this story and can answer it."

(iv) *General conclusion from Example I*

The speed and confidence with which all subjects made their predictions, the fact that these predictions ran along easy, stereotyped lines, and the high percentage of agreement, all indicate that the presentation of the material immediately caused reference to be made to certain widely accepted generalizations.

It is apparent, especially from the forceful manner in which they are explicitly formulated, that the belief in these generalizations, and their utilization in the present instance at least, are accompanied, and probably to a large extent determined, by strong emotional factors. This is most clearly seen when their implied contradiction arouses actual resentment.

Once these generalizations have been accepted as relevant to the presented material, they have such 'weight' as almost completely to dominate all subsequent 'thinking', and to exclude careful consideration of the trend of the story and the significance of the data which it contains. This is perhaps most clearly seen in the last paragraph, when the generalizations evoked by the mere mention of Joab's sorry plight are such as to make practically all subjects overlook the fact that this same sorry plight makes it all the easier for the hunters to carry out their professed intentions.

V. Example II

(i) *Material*

Tunersuk

"A month ago", said the young hunter, "Tunersuk went away in his *kayak*: none of us have ever seen him again." "Perhaps he had an accident", said the trader. "That is not likely", said the young hunter; "he was the best *kayak* man in the whole

settlement." "He certainly was a fine hunter", said the trader, "but he would not fish or take sharks, and so lost much money." "Did you not often advise him to fish?" asked the catechist. "Oh, yes", said the trader, "but he always replied that fishing was work only fitting for women and boys. He was proud." "I should say that he was just lazy", replied the catechist. The young hunter looked angry: "No", he said, "he was not lazy: he thought it best to live like the old Greenlanders: and he used to say that it would be better for Greenland if the white men had never come." "Nonsense!" said the catechist: "where would we all be now without the white men?" "*You* certainly would not be here", replied the hunter.

The trader spoke again: "It is foolish to quarrel about things which cannot be changed. These last two years Tunersuk suffered every kind of misfortune—he lost his wife, his dogs, his sledge and his *kayak*. Then he himself fell ill: I think that he knew that he would not live long. Then he began to hate the white men and to blame them for everything. I offered him a boat and tackle to take sharks, but he angrily refused; and when I gave him money he threw it back at me. After that I could do nothing more."

"He came to me before he left", said the young hunter, "I was always his friend. He told me that he was going away for ever because there was nothing more for him to do here." "I wonder what has happened to him", said the trader. "I believe", said the young hunter, "that he is now a *kivitok*.[1]"

The catechist laughed, but rather uneasily.

The questions in this instance are:

(1) What do you think happened to Tunersuk?

(2) Which of the three men, trader, hunter, and catechist, do you like *most*?

Which do you like *least*?

(3) What do you think of Tunersuk's behaviour and opinions?

(ii) *Method of presentation*

This story was presented to 88 subjects—24 in Igdlorssuit, 21 in Qaersut and 43 in Umanak. Of the total number, 47 were men, 29 women and 12 children under 14 years of age. All results were obtained from group experiments conducted as described above (Example I, p. 298).

This material differs from Examples I, III, IV in demanding a *retrospective* rather than a *predictive* judgement. It will be seen, however, that this difference is not so essential as to affect the general manner in which subjects deal with the data presented. The 'supernatural element' was introduced because it is maintained that a belief in the traditional Eskimo 'demonology' persists even among the more enlightened West Greenlanders, and can affect both their thinking and their behaviour.

[1] A man who, having fled the settlements, maintains himself in a demoniac state in the mountains.

(iii) *Analysis of results*

Table IV classifies and shows the relative emphasis on the answers to question 1, "What happened to Tunersuk?"

Table IV

	Igdlorssuit			Qaersut			Umanak			Total	
		No.	%		No.	%		No.	%	No.	%
Became *kivitok*		6	25		6	28		29	67	41	47
Did not become *kivitok*:											
Perished in *kayak*	14			12			12				
Insane (but *not kivitok*)	1	16	67	1	14	67	1	14	33	44	50
Stayed away, alive and well	1			1			1				
Doubtful		2	8		1	5		—	—	3	3
Total		24			21			43		88	

The distribution in the two hunting settlements is very similar, about 25% of each group saying that Tunersuk turned *kivitok*. It is interesting that in the more 'civilized' community of Umanak this percentage should rise to 67–55% of the men, 64% of the women and nearly 100% of the children. On an over-all count, nearly 50% give a 'superstitious' explanation of Tunersuk's disappearance; but if subjects under 14 years of age are eliminated (as they probably ought to be, for they occur only in the Umanak group), then the percentage of 'supernaturalists' is 40%, of 'rationalists' 60%.

The division of opinion was very marked, practically all the 'supernaturalists' saying that Tunersuk *certainly* became *kivitok*, and all the 'rationalists' that he *certainly* did not. Many of the more enlightened and intelligent Greenlanders are most emphatic supernaturalists, e.g. the 'catechist' (native teacher and lay preacher) of Qaersut, who says: "Tunersuk certainly became *kivitok*, after the old Greenland custom." The headman of Igdlorssuit prefers to suspend judgement: "This is the old Greenland custom", he says, "I can well believe that a human being can turn *kivitok*, though I have never seen any such thing." One of his brothers, on the other hand, states that, "no one can become *kivitok* in *our* time", in contrast to yet another brother, who, along with four Igdlorssuit and five Umanak people, believes that Tunersuk must have turned *kivitok* "because he *said* that he was going to, and *could do so if he wished*".

Questions 2 and 3 concern only the subsidiary aims of these experiments. Table V brings out, with small variations, for each settlement, a fairly definite order of preference.

Reasons for preference are seldom given. They do appear, however, among those who, like the hunter, are Greenlanders of the 'old-fashioned',

independent type. One of them says, "It *is* better to hunt than to fish, and there are still enough seals for all men to be hunters"; and another (from Umanak) believes that "all hunters have a mind of their own"—perhaps a subtle criticism of his fellow Greenlanders in the employment of the Danes! So few subjects answered the second part of question 2 that their answers are not worth tabulating.

Table V

Which of the three men do you like most?	Idglorssuit (24)	Qaersut (21)	Umanak (43)	Total (88)
Trader	7	11	19	37
Catechist	9	1	13	23
Hunter	2	7	11	20
No preference expressed	6	2	—	8

Table VI classifies the types of answers received to question 3.

Table VI

What do you think of Tunersuk's behaviour and opinions?	Igdlorssuit (24)	Qaersut (21)	Umanak (43)	Total (88)
Good	8	5	8	21
Bad	7	9	31	47
First good, then bad	1	—	2	3
Not evil, but mad	1	—	—	1
Indefinite answers	1	4	—	5
No answer	6	3	2	11

No justification of Tunersuk's behaviour as good was attempted, though several subjects considered him good "in spite of everything". Many reasons were given for condemning him as bad; "he turned *kivitok*"; he was lazy, improvident, did not live lawfully, was not good-hearted, was too easily offended and so on. The only explicit reference to Tunersuk's hatred of the white man is made by the headman of Igdlorssuit and his brothers, who find it a fault in Tunersuk that "he was not friendly to the white men, who are our helpers".

(iv) *General conclusions from Example II*

This story aroused no criticism of the type levelled against Example I (p. 300, above). On the contrary, in all three settlements the Greenlanders said: "Of all the stories which you have told us, Tunersuk is the best, for it describes the kind of thing which might happen to us Greenlanders at any time." At the mention of *kivitok*, and occasionally while subjects were recording their answers, there was some laughter, but it was excited and semi-apologetic rather than scornful, and therefore there is good reason to assume that in general the emotional reaction to the story was such as to encourage sincerity.

The presentation of the material evidently suggested immediately two sharply contrasted types of explanation. Of the two, the 'supernatural', probably on account of its emotional associations, has by far the greater 'weight', for although it is *accepted* by less than half the number of subjects, it is *mentioned* and *expressly rejected* by nearly all the others.

'Thinking about' the presented situation did not entail any very careful consideration of specific data, but resolved itself into a question of accepting or rejecting a conventionalized and formerly universally accepted type of explanation. Acceptance and rejection, as shown by the speed and confidence with which subjects made their decision, and by the kind of reasons occasionally expressed (p. 302, above), seem to have been largely pre-determined by factors quite removed from the presented situation; and where actual reference is made to the text, the acceptance of the *kivitok* hypothesis causes ten subjects to twist Tunersuk's words to suit their case (p. 302).

It appears, therefore, that in spite of the one demanding a predictive and the other a retrospective judgement, the subjects' modes of dealing with both Example I and Example II display much the same general characteristics.

The value of the answers to questions 2 and 3 will be considered later in a general discussion of the secondary aim of these experiments.

VI. Example III

(i) *Material*

The Famine

Many years ago ten Englishmen came to Greenland on an expedition. They brought much food and equipment, because they expected to work for two years. The place where they built their house was very lonely and difficult of access, and the only people there were some twenty Greenlanders who lived in a small settlement $\frac{1}{2}$ km. north of the white men's house. These Greenlanders told the Englishmen that sometimes in the winter they were for long periods cut off from all communication with other settlements.

The first year all went well. The Greenlanders had good hunting, the white men collected rocks and flowers; they flew balloons; they dug up old houses and in general did all the peculiar things which white men do, and seemed highly pleased with themselves and their work. Frequently they asked the Greenlanders to help them in one way or another, and the Greenlanders were always willing to do so. The Englishmen paid good wages and gave many presents; and they always made the Greenlanders welcome in their house. And when the Greenlanders asked them to a 'dancé-mik' or a 'café-mik' they always seemed glad to come. Of course the Greenlanders,

being poor, could not give such fine presents as the Englishmen, but none the less many of them gave small gifts which the Englishmen seemed to prize highly.

One day, after a Greenlander called Jakob had brought such a gift and gone away, one of the Englishmen said to his companions: "You know, I believe that Jakob, and most of the Greenlanders, give us these gifts only because they hope to get a better one in return." "That's a mean thing to say", replied the man who had received the gift, "and I'm sure that it's untrue and unfair to Jakob and all our Greenland friends." "You are very generous, but a little simple-minded", said the first ironically. "For my part I don't believe that white men and Greenlanders can ever understand each other, or be real friends. The Greenlanders do not like us *for ourselves*, for *what we are*, but for what they can get out of us; and if we'd nothing to give these people here then there would be a quick end to this so-called friendship of ours." "You are wrong", answered the second man, hotly. "After all, white men and Greenlanders are human beings, made after the same pattern, and living much the same sort of life—eating and sleeping, loving and hating, crying and laughing, growing up, growing old, and all meeting the same fate in the end. And in these things are found the beginnings of true friendship. I am sure", he added, "that Jakob and I are real friends; that he likes me as I like him, and that he would stand by me in any danger." "Perhaps; perhaps not", answered the other. "I'd like to see what would happen if it was a question of our lives or the lives of our Greenland friends. I bet you we'd be left to shift for ourselves."

The very next winter his words were put to the test. The early months were wild and stormy, and little hunting was possible either on sea or land. The new ice was broken over and over again, and everywhere in Greenland the people suffered much from cold and hunger. The Greenlanders in the little lonely settlement beside the white men suffered most of all. They had no trading-store; they could not travel to any other settlement for help, and in the end it was the white men who saved them. They shared with them the food they had brought from Europe, giving to every one that was in need, and taking care above all that no little child should ever go hungry.

At last the ice came, and the Greenlanders began to get a few seals—but not many, scarcely enough to feed themselves. The white men could not hunt, and one day they came to the Greenlanders and said: "Our food is nearly finished because we shared it with you when you had none. Can you help us now?" The Greenlanders did not know what to do. "We ourselves", they said, "have just enough from day to day, and no one knows but that to-morrow we may have nothing. Already some of us are weak and ill from lack of food." When the leader of the Englishmen heard this, he said, "Do what you think best. We shall return in three days' time to hear your decision."

They went away, and the Greenlanders began to discuss the matter among themselves. "We ought to help them because they helped us", said some. "Perhaps", said others. "But you forget that when they helped us they had plenty, and both they and we confidently expected that good hunting would come soon. Good hunting has not come. If we share our food with them now then it is highly probable that both they and we shall perish. But if we keep what we have, then we may live. It is better that some should live than that all should die; and it is natural that we should save ourselves."

"It may be *natural* to save ourselves", said the catechist at this point, "but it would be cowardly and ungrateful and dishonourable. You remember what the Bible

says—'Greater love than this hath no man, that he lay down his life for his friend'.'' "Yes. For his *friend*, for a man of his own race and blood!" said a young Greenlander. "But these Englishmen are strangers to us." "Let us share what we have with them. Hunting may improve", said one of the women, hopefully. "There will be no good hunting this winter", replied an old hunter.

And so they talked on, and eventually all were agreed as to what it was best to do. And on the third day the Englishmen came to hear their decision.

The questions in this instance are:

(1) What did the Greenlanders decide to do?

(2) What do you think it will be best to do, and *why* will it be best?

(3) Do you think Greenlanders and Europeans can understand each other and be real friends?

This is by far the longest story used in Greenland. It was written mainly in order to see whether any real attempt would be made to sift a very complex set of data. The risk of including some rather critical statements about Greenlanders was taken because it was thought that, in such a setting, they would not be likely to cause offence, and might even encourage a free and genuine expression of opinion.

(ii) *Method of presentation*

This story was used in Umanak only, with a group of thirty-four subjects—sixteen men and eighteen women. The method of group presentation described above (p. 298) was adopted.

(iii) *Analysis of results*

Tables VII and VIII show the kind of answer received to questions 1 and 2, and the relative emphasis on each.

Table VII

What did the Greenlanders decide to do?	Men (16)	Women (18)	Total (34)
To help the Englishmen	15	17	32
Doubtful	1	1	2

Table VIII

What do *you* think it will be best to do, and *why*?	Men (16)	Women (18)	Total (34)
(*a*) To be friends and give mutual help	13	16	29
The Greenlanders ought not to help *too much*	1	—	1
No answer	2	2	4
(*b*) The Greenlanders must help, because they have received help	9	13	22
For the sake of the good name of all Greenlanders	—	2	2
No answer	7	3	10

No one predicts that the Greenlanders will *not* help, though some say that "they will give only what they have to spare, and the Englishmen

will understand that". Only two subjects, both old hunters, refer to the arguments in the text *against* helping, and they merely dismiss them without discussion as "stupid and wrong". Question 1, in fact, was answered with the same speed and confidence as was apparent in subjects' reactions to Examples I and II, and practically all answers seem to have been immediately and completely dominated and directed by the sentence, "we ought to help them, because they helped us", for there is an almost monotonous recurrence of the statement, "The Greenlanders must help the Englishmen, because they helped them", along with an apparent blindness to obvious practical difficulties, and the force of any other argument whatsoever. One can only speculate, of course, as to *why* this phrase should be selected and dominant, but at least part of the reason must be that this particular story succeeded in provoking the same 'group self-consciousness' and 'group pride' as were apparent in some subjects' reactions to Example I (p. 300). This emotional element, aroused evidently by the preliminary argument about the giving of gifts and the whole trend of the narrative, is reflected in the tone and phrasing of numerous generalizations, e.g. "Greenlanders are more backward in their way of thinking than other folk, but they gladly give something on all occasions"; "Greenlanders are exceedingly helpful"; "we Greenlanders are in the habit of repaying what has been done to us"; "we Greenlanders have a custom from old times of helping one another."

Question 2 was put in order to see whether any subjects would differentiate between what *would* be done and what *ought* to be done. As one would expect from the general attitude adopted towards question 1, and the type of answer which it elicited, no such distinction is drawn. The unusual hesitancy which was shown in dealing with question 2 can therefore only be attributed to the fact that most subjects could see no real difference between it and question 1, and were wondering why they had been asked to repeat themselves.

While the men were nearly all agreed that it is best to help, they are more reluctant than the women to say *why* they think so. The only explicit justification of giving help is a repetition of what was said in reply to question 1: "The Greenlanders must help, because they have received help". To this two women add, "and also for our own sake", meaning evidently, "for the sake of the good name of all Greenlanders".

This kind of 'moral justification' is interesting in itself, but the fact that no attempt is made to consider in detail practical working difficulties is rather disappointing. The only suggestion of this kind was made in conversation by a Dane, who said that the Greenlanders would probably

help the Englishmen, but in dividing the spoil would give priority to their own hunters, who at all costs must be sufficiently nourished to hunt. This was the old Eskimo custom, and perhaps might be referred to if ever this material is used on some future occasion in isolated hunting communities.

Question 3 was answered as shown in Table IX.

Table IX

Do you think Greenlanders and Europeans can understand each other and be real friends?	Men (16)	Women (18)	Total (34)
Yes. Both	9	16	25
Good friends, but no real understanding	3	—	3
Understanding, but no real liking	1	1	2
No. Neither	1	—	1
Doubtful	2	1	3

The women are much readier than the men to give an unqualified 'yes'. This is further emphasized by a fact not shown in Table IX, that five of the nine men make this reservation—that Greenlanders and Europeans will only understand each other and be real friends *if* they live together for a long time, and the education of the Greenlanders improves. Two of the women mention the former of these two conditions, and two others, that it will be necessary to learn a common language. An interesting expression of group attitude is found in the statement, made by two subjects (a man and a woman), that real understanding and real friendship are possible "in spite of the great difference of wealth, and although Europeans are wiser". On the other hand, while the three men who think that Greenlanders and Europeans can only *like* each other, acknowledge this "difference in intelligence", they find it, and also "difference in custom", an insuperable obstacle to real understanding.

(iv) *General conclusions from Example III*

On presentation of this material, one item was immediately picked out and so heavily 'weighted' as to be completely dominant.

Reasons why this should be so have been suggested (p. 307, and cf. Example I, p. 300), and its effects are, as in Example I, the production of easy, stereotyped answers and the evoking of a distinctive type of generalization.

The ease with which question 1 was answered, as compared with question 2, the failure to differentiate between them, and the marked absence of any discussion of practical difficulties, all indicate a tendency

on the part of most subjects to identify what would happen with what they should wish to happen.

Question 3 will be considered later in a general discussion of the secondary aim of these experiments.

VII. EXAMPLE IV

(i) *Material*

The Haunch of Venison

It was unfortunate for Johan that the dogs should have passed at that moment. You see, Johan had been working for the outpost trader for nearly six months, and on the whole he and the trader understood each other fairly well. Johan was capable, and the *bestyrer*,[1] who was a careful man himself, liked him for that, and also because he was clean and quiet. The trader, like most men who talk a good deal, was rather hot-tempered, but as Johan knew, his bark was worse than his bite, and he was really rather a kind-hearted man. So Johan kept out of his way when he was in a bad mood, and generally the two of them got on very well together. But Johan had one fault. He himself could not see anything wrong in what he did, so perhaps one cannot blame him too much. After all, the trader had more than he required of most things, and if Johan *did* help himself now and again to a little food, a little coffee or a few cigarettes —well, as long as the trader knew nothing about it, then he was no less happy and Johan was a good deal happier.

At the end of his fourth month, and again during his fifth, the trader told Johan in no uncertain language what he would do to anyone that he caught pilfering. He did not accuse Johan directly, but Johan was wise enough to take the hint, and mend his ways.

And that brings us back to the dogs. Johan was one day taking a haunch of venison out to the little cold store-house where the trader kept his meat. Johan happened to forget the key of this little house, which was always kept locked. The venison being heavy, he hung it on the handle of the door and ran back quickly for the key. Just then the dogs came round the corner, and by the time Johan got back there wasn't much venison left. First of all Johan thought it would be best to tell the trader what had happened; but then he remembered that the trader was busy, and in a bad mood. So Johan said nothing about the accident, and hoped that the trader would forget his venison.

A few days later, however, the trader asked for his venison, and Johan had to tell him how he had lost it. Poor Johan did not tell his story very well, or very convincingly, for he saw from the very beginning that the trader was not believing him. When the story was finished, the trader laughed—a short hard laugh—and then he looked at Johan and said: "You are the only man other than myself who is allowed to use the key to my storehouse. I don't think it was the *dogs* who ate the venison." He laughed again, but only with his mouth, and Johan could see the anger growing in his eyes.

[1] Trader.

The questions in this instance are:

(1) What will the trader do?
(2) What will Johan do?
(3) With whom do you sympathize more, and why?

(ii) *Method of presentation*

This material was used in Umanak only, in a group experiment with thirty-three subjects—sixteen men and seventeen women. The method of group presentation described above (p. 298) was adopted.

From the general type of reaction observed, it seems that this attempt to write a story which would be taken light-heartedly was fairly successful. All laughed expectantly at the title, as if it gave promise of something amusing to follow. There was little or no reaction, amused or otherwise, to the account of Johan's pilfering,[1] but the loss of the venison, and Johan's subsequent behaviour, up to the telling of his story to the trader, caused hearty and prolonged laughter. The trader's reply, on the other hand, was not considered funny, but had a marked sobering effect on the group as a whole.

(iii) *Analysis of results*

Tables X–XII show the kind of answer received to questions 1–3 respectively. Questions 1 and 3 were answered promptly; question 2 much more deliberately.

Table X

What will the trader do?	Men (16)	Women (17)	Total (33)
Dismiss Johan	5	3	8
Be extremely angry and scold Johan	7	12	19
Be extremely angry, and then pity, understand and pardon Johan	3	2	5
So angry at losing his venison that he will not eat for a long time!	1	—	1

Table XI

What will Johan do?	Men (16)	Women (17)	Total (33)
Accept his dismissal and/or give up his job	5	10	15
Give up his job, but be taken back	1	—	1
Explain, apologize	10	5	15
Become even more zealous	—	2	2

[1] This kind of petty pilfering and the fact that many Greenlanders condone it, even if they do not practise it, is said to be a relic of the old Eskimo communism. The Greenlanders know very well how much the Danes detest it.

Table XII

With whom do you sympathize more?	Men (16)	Women (17)	Total (33)
Johan	14	16	30
Trader	1	—	1
Neither	1	1	2

Although the total number of subjects is small, the following comparative percentages are interesting, and seem to indicate significant differences of opinion. Taking all the answers together, approximately 54% predict only a short burst of anger on the part of the trader; 25% are certain that he will sack Johan; and 15% think that he will be reasonable, and compassionate. Opinion is almost equally divided that Johan will either give up his job or explain, apologize and presumably continue work.

Taking men and women separately, however, it is clear that the women are more inclined than the men to predict a short burst of anger on the part of the trader, and rather less inclined to think that he will sack Johan; but nearly 60% of them think that Johan will *voluntarily* give up his job, and only 30% that he will try to explain. In the case of the men, these last two percentages are exactly reversed.

The opinion expressed by the women is in keeping with what the Danes say of their Greenland *women* servants—that you can scarcely scold them for fear they run away! In this instance, therefore, the women's answers may have been largely determined by their imputing to Johan conduct more likely to be expected of a native woman. The opinion expressed by the men, on the other hand, is in keeping with the treatment they are accustomed to either as casual or regular workers under the Danes. Even in the coal-mine and marble quarries, when they are in regular employment for wages, conditions make it inconvenient to have them idle, and a kindly administration has so far discouraged any attempt to introduce strict working regulations, and the threat of dismissal and unemployment. Greenland men, therefore, have little experience or fear of being sacked, and perhaps that is why the majority of men subjects in this experiment seem confident that a full explanation and apology will be effective.

Sympathy is overwhelmingly in favour of Johan: "because he'd been unlucky in losing the venison, and in being falsely accused" (fifteen subjects); and "because he was afraid and couldn't tell his story properly" (four subjects). Four other subjects pity him for his faults; "he'd committed crimes"; "he couldn't go straight"; "he was foolish and negligent"! The one man who liked neither the trader nor Johan

thinks that the former is "hot-heated and stingy" and the latter "dishonest"; while the humourist who predicts that the trader will not eat for a long time out of sheer annoyance, sympathizes with him greatly, but thinks that "he'll duly recover his appetite and have Johan back"!

(iv) *General conclusions from Example IV*

The general types of answer to question 1 are so much in keeping with the description of the trader's character in the text that it seems reasonable to infer that this description was heavily weighted by most subjects, though no one of these makes explicit reference to it. In other words, subjects seem in this instance to have utilized specific data presented in the text, and not to have relied to any appreciable extent on socially accepted generalizations concerning white men and their behaviour.

In answering question 2, on the other hand, both men and women subjects go outside the limits of the data presented, the men predicting that Johan will behave as most Greenland *men* would in similar circumstances, and the women that he will behave as most Greenland *women* would.

Although no explicit generalizations are made, the answers (especially to question 2) seem to reflect definite and fairly widely accepted opinions concerning the reciprocal relationship of Danes and Greenlanders, and its expression in practice.

VIII. Results

(i) *The nature of constructive thinking*

In preliminary and tentative experiments of this kind one is more concerned with the value and possibilities of the method employed than with reaching immediately any very definite conclusions. None the less, the four experiments described, and others conducted last summer in Greenland, have served to draw attention to certain features which seem characteristic of all subjects' reactions to this type of problem situation. These characteristic features are as follows:

(i) A quick confident response.
(ii) Easy, stereotyped answers.
(iii) Either a high percentage of agreement, or disagreement along two or more very definite lines.
(iv) The frequent formulation of generalizations of a very distinctive type.

(v) A constant tendency to pick out one item, or set of items, to which completely dominating weight is given, and very rarely any attempt to utilize all the data presented.

(vi) A tendency to identify what *will* happen with what 'must', 'ought', or 'should' happen.

(vii) Consequent upon (v) and (vi), a tendency to make the facts fit the solution, rather than vice versa.

(viii) In (iv), (v), (vi) and (vii), considerable evidence of emotional involvement.

What the experiments demonstrate most clearly is that constructive thinking about problem situations of this type is highly 'conventionalized". This is neither an unexpected nor a novel conclusion, and any attempt at this stage further to define the precise nature of this 'conventionalization' would be speculative and premature. But the manner in which the fact of conventionalization has been demonstrated serves to direct attention to another fact which admits of direct investigation by objective methods, viz. the fact of *selection*. By the comparative study of social groups within one culture and of different cultural groups, it should be possible to obtain by more carefully controlled experiments valuable evidence as to how this process of selection operates, and some indication of the relation in which it stands to conventionalized patterns of thinking.

One quite definite result can, however, be immediately stated. A series of experiments, similar to those described, but using problem situations constructed to suit a different social background, has been carried out in Cambridge by Prof. F. C. Bartlett.[1] A comparison of notes and results proves at once that the kind of reaction observed in England, and the type of answer received, display precisely the same general characteristics as those described above. It therefore seems reasonable to infer that the mental processes involved in dealing with this kind of problem, at least, are, in each instance, *psychologically* identical, and that the contention that civilized persons strive always to reach a well-balanced coherent conclusion, while the primitive is willing to tolerate incoherence and even 'considerable contradictions and contrasts' is, if not unfair to the primitive, at least highly flattering to ourselves.

(ii) *The method*

(*a*) *In relation to the primary aim of the experiments.* From the examples given, it will be seen that the method possesses the advantages

[1] The results of these have not yet been fully published.

claimed for it above (p. 297), and also, of course, that it can be extended and improved. As the tendency to conventionalize, and to go beyond the limits of the data presented, must be at least partly due to the type of 'incomplete situation' employed, the first step is to devise and try out 'control situations' which can be solved, to a greater or less degree, within their own limits.

(*b*) *In relation to the secondary aim of the experiments.* As an enquiry by this method into the attitude of Greenlanders to Europeans would require, in order to yield results of any value whatsoever, a much more extended survey than was possible last summer, discussion of the secondary aim of the experiments is necessarily confined to a few notes on the use of this method with primitive people.

As compared with the questionnaire, this method has the following advantages:

(1) Nearly all primitive people, and certainly all Greenlanders, in response to a direct question tend to give an answer which they think will please the questioner. In all the experiments conducted in Greenland, no evidence of this kind of insincerity appears, and there is considerable evidence pointing in most instances to genuine and sincere answers.

(2) Insincerity seems to have been eliminated, and sincerity encouraged by (*a*) the general group attitude to *any* story, and (*b*) the specific emotional 'atmosphere' created by the actual kind of story used.

(3) Within the context of a well-chosen story, points which might otherwise cause offence can be safely raised.

(4) By referring them to an actual concrete situation, the questioner can be reasonably sure that his questions have the same essential meaning for his subjects as they have for himself.

(5) If we ask an individual a set of isolated questions concerning group relationships, then we can scarcely hope to know what aspects of his experience he has recourse to in framing his answers. If, however, we describe a situation in which certain given factors are *possibly* relevant, then we can observe what selection and weighting of these factors occur, and what further factors, if any, are introduced. In other words, we are enabled partially to control our subjects' 'frames of reference'.

Too much indirectness, on the other hand, is not desirable. Thus in Example II (Tables V and VI) and again in Example IV, Table XII, we have sets of answers which *probably* reflect fairly definite and widely accepted opinions concerning the reciprocal relationship of white men and Greenlanders, and its expression in practice, but which leave the

exact nature of those opinions more or less a matter of speculation. Although it is doubtful whether Greenlanders (or indeed *any* social groups) could give full and precise *verbal* expression to such opinions, none the less a partial expression might be obtained by fairly direct questions following upon the presentation of a relevant situation (such as question 3, Example III), and by employing the method of individual presentation and oral response (Example I) which, as we saw, gave by far the richest crop of significant generalizations.

(iii) The method is practically self-maintaining, producing unsolicited generalizations which, if thought significant, can be embodied in further specially devised situations. The study of the nature, frequency and distribution of the conventionalized opinion of which these generalizations are a partial expression can thus be extended indefinitely.

REFERENCES

(1) LÉVY-BRUHL, L. (1926). *How Natives Think.*
(2) SCHULTZ-LORENZEN, C. W. (1926). *Greenland*, vol. II.
(3) BARTLETT, F. C. (1938). "The co-operation of social groups." *Occupational Psychology*, XII, 30.
(4) CARMICHAEL, D. M. (1938). "The co-operation of social groups." *Brit. J. Psychol.* XXIX, 206, 329.

(*Manuscript received* 10 *April* 1939)

THE CONCEPT OF THE 'SCHEMA'.[1]

By MARY L. NORTHWAY

From the University of Toronto

PART I

I. *Introduction* (p. 316).

II. *The concept of the 'schema' and its manifold uses* (pp. 317–318).

III. *The immediate background from which Bartlett's theory developed* (pp. 318–323).

IV. *Purpose and general method of the experimental investigation* (p. 323).

V. *Preliminary experiments* (pp. 323–325).

(i) *The influence of the 'difficulty' of the material on remembering* (pp. 323–324)

(ii) *The influence of age and social group on children's remembering* (pp. 324–325).

I. Introduction

The purpose of the experiments herein reported was to examine the concept of the 'schema' used by Bartlett in his theory of remembering.[2] The various meanings he gives the term and the sources from which it developed are examined. Investigations in primary schools pointed to an interpretation which appeared to be sufficiently definite to direct research, and which serves to indicate some educational implications of Bartlett's point of view.

The work of Bartlett has had special significance in two ways. First, it has demonstrated that a developmental psychology, of the type propounded by Ward, may be subjected to rigid experimental investigation. Secondly, it has reaffirmed the notion that memory is no passive storehouse in which items are retained and arranged according to certain mechanical principles, but rather is that process by which the active individual recreates his past experience in ever-evolving forms. This dynamic process, remembering, is influenced by the individual's "appetites, interests, ideals and social background"; and of particular pertinence in the organization and re-organization of experience is the notion which Bartlett has termed the 'schema'.

[1] Abstract of a thesis accepted by the University of Toronto for the Degree of Doctor of Philosophy.

[2] References, unless otherwise indicated, refer to Bartlett's *Remembering* (1). All references will be given at the end of Part II of this paper, to be published in a subsequent number of the *British Journal of Psychology*.

II. THE CONCEPT OF THE 'SCHEMA' AND ITS MANIFOLD USES

In the study of remembering as an active process Bartlett attempts to give the concept of the schema an important place. Exactly what this place is, is difficult to define, not because the term is used vaguely, but because it is used in so many different ways. And because one use is interchanged with another without any explanation, a great deal of confusion arises.

Four different uses of the term are listed below:

(1) The schemata are considered to be the *forces* within the individual which determine what the subject's reconstruction will be:

> Determination by schemata is the most fundamental of all ways by which we can be influenced by reactions and experiences which occurred sometime in the past (p. 201).
>
> What, precisely, does the 'schema' do? Together with the preceding incoming impulse it renders a specific reaction possible. It is, therefore, producing an orientation of the organism towards whatever it is directed to at the moment. But that orientation must be dominated by the immediately preceding reaction or experience. To break away from that the 'schema' must become not merely something that works the organism, but something with which the organism can work. The organism discovers how to turn around on its own 'schemata'. In other words, it becomes conscious (p. 207).

(2) The schemata are the *forms* in which the individual preserves material.

> It looks as if that preservation of material which is required in recognizing is normally a preservation of schemes, of general settings, of order or form of arrangement (p. 195).

(3) The schemata are 'storehouses' in which *content* is retained while it is being reorganized.

> In fact this is one of the great functions of images in mental life: to pick items out of the 'schemata', and rid the organism of over-determination by the last preceding member of a given series (p. 209).
>
> All of us in reference to some of our 'schemata' have probably completed the model and now merely maintain it by repetition (p. 203).

(4) 'Schema' is used in a way approximating the notion of the *apperceptive mass*.

> I think probably the term 'organized setting' approximates most closely to the notion required (p. 201).
>
> The influence of the 'schemata' is influence by the past. In its schematic form the past operates en masse, or, not strictly en masse because the latest incoming constituents which go to build up a 'schema' have a predominant influence (p. 202).

Although this analysis separates the various meanings of the term artificially and implies they are more distinct, one from another, than they actually are in Bartlett's use of them, the confusion resulting from using so many conflicting meanings is apparent. It should be noted that although Bartlett states that he is considering remembering as an active re-organizing process, in many of the cases he gives, the idea of memory as a 'storehouse'—as content carried forward—seeps in all too easily. In order to clarify and appreciate the various connotations given the term, the first piece of research was to trace the sources from which it came. This consisted of a study of the immediate background from which Bartlett's work developed.

III. The immediate background from which Bartlett's theory developed

The problem of this section is to select some of the influences which have directed Bartlett's psychology and his concept of 'schemata' as he uses it in his theory of remembering, with the purpose of using this information to clarify the meanings he gives to the concept. While the notion could be traced back almost endlessly, this section is limited to the discussion of its immediate background.

If the title of Bartlett's book, *Remembering: A Study in Experimental and Social Psychology*, is analysed, the three main influences on his thought are revealed. The 'experimental' aspect is closely associated with the work of Sir Henry Head and the physiological group; the 'social' aspect is related to the influence of Rivers and the Cambridge anthropological interest, which developed from the days of the Torres Straits expedition; and the 'psychology' is found to be in direct descent from that of James Ward, the most influential of Cambridge teachers in Bartlett's time.

Although these three influences have been isolated for discussion, in order to appreciate Bartlett's psychology it is necessary to bear in mind the close interrelationships which exist between various fields of study in the Cambridge setting. Demarcation between fields of research, and departmentalizing of subjects, are far less rigid than in Universities on the American continent. Facts, theories and investigations from the various sciences are absorbed together, and a psychology to endure must not only be consistent in its own right, but must not conflict with findings recorded by the physiologist and sociologist. Bartlett's knowledge of, and keen interest in, these fields, and his own background in the

'moral sciences', influence very greatly the psychology he derives, and may be responsible not only for the breadth of his point of view but for some of the confusions which arise. This is particularly true with regard to the notion of schemata, which Bartlett derives from psychology proper, from physiology and from anthropology; and it is in attempting to form a concept which is in harmony with thought in all these fields that a great many of the difficulties arise.

(*a*) *Sir Henry Head.* In order to appreciate Head's influence on Bartlett the fact that a relationship more personal than purely academic existed between the two men should be taken into account. Each in the introduction to his major work (1), (8) acknowledges his debt to the other, and Bartlett in a critical notice of Head's two volumes on Aphasia writes of it in eulogistic terms which show a close sympathy for its development and implications.

> For the psychologist these are two of the most important books which have appeared for many years. They are scientifically and aesthetically satisfying....I believe these books will have a permanently important place...in the history of shaping psychology into a science...the general viewpoint will remain of abiding interest and lasting importance. And as with all genuinely valuable achievement, Dr Head's life work and particularly this crowning effort will be a most powerful stimulus for further research (2).

Bartlett, being so much impressed with Head's work, felt very strongly that his own must not only be in harmony with it, but also indicate the close relationship by using (in modified form) one of the terms which Head had made so significant in recent biological thought. This term was 'schema'. Head defined it as:

> That combined standard against which all subsequent changes in posture are registered before they enter consciousness (8).

Although Bartlett in taking this term over from Head claims his right as a psychologist to modify it to mean "an active organization of past reactions" (p. 201), the above definition, with its emphasis upon a 'standard', contains the idea which appears, here and there, in Bartlett's work—that the schema represents the combined *content* of previous reactions heaped up like the *apperceptive mass* and taking on a certain *form*. These three uses of the term have at least one of their sources in the influence of Sir Henry Head. The failure of Bartlett to retain the modification of the schema which he has stated, and his regression to its use as a somewhat static standard, constitute one reason for confusion in its meaning; the unsuccessful attempt to reconcile this concept of a standard with a Wardian psychology is another.

(*b*) *James Ward.* The principles of Ward's psychology are implicit throughout the work of Bartlett, who was not only his pupil but who maintained a close relationship both academic and personal with him for many years. Ward's psychology has been summarized by Brett in this way:

> The unity of consciousness is a continuance; the whole process must be regarded as the gradual differentiation of a primary unity...the mind is an organism which grows by continuous development and differentiation into a collection of parts that develop in sustained relations and never change without involving reciprocal change through the whole structure (6).

These ideas of continuance and development are assimilated so intimately by Bartlett that he assumes them throughout his writing without making any explicit reference to them. It is in a psychology based on these that schemata in remembering have to be made meaningful.

While Bartlett's general point of view is based on Ward's psychology, Ward had also a specific influence on Bartlett's theory of remembering. Writing an obituary of Ward in the *American Journal*, Bartlett states:

> Another part of psychology had an irresistible attraction for him, the study of memory. Here if anywhere was to be found actual evidence of the integrity and independence of mind, and the indissoluble character of individual experience (3).

Ward, who himself emphasized the influence of Lotze in his thought (3), acted as a direct bearer of the idea of the activity of memory from Lotze to Bartlett. This legacy included with it the concept of the schema thought of as a *plan* of activity, the *form* which the activity will take. Brett writes:

> To Lotze and all idealistic psychologists the memory is the crucial instance; there is no faculty of memory, no cerebral organ, no 'storehouse of ideas'; only the living continuity of the soul and its power of reproducing its previous activities....Memory does not keep any picture; what it really retains is a kind of *schema*, a plan of action (6).

This is the meaning schema has in the psychology Bartlett assumes, but it is a meaning he does not make explicit or develop. It should be kept in mind as a possible interpretation of the term when it is used in a theory which considers remembering as an active process belonging to a 'developmental' psychology. This use will be referred to in the experimental section of this paper.

The most important influence Ward has on Bartlett's theory of remembering, however, is the significance he gives to interest. Ward

casts aside association as an explanatory principle and substitutes continuity for contiguity (11). The continuity of the individual's activity is considered the integrating principle. In answer to the question of how new wholes are integrated, Ward restates the Lotzian point of view "that movements of attention determine what wholes will be formed and these are based primarily on *interest*" (6). Ward does not think of items of content being absorbed and then integrated by some superimposed process of association, but rather, he considers the activity of the individual through his interest to be in its very nature integrating and developing.

This function of interest is taken over and becomes important in Bartlett's theory. The schemata are built up by several processes in human life; but the chief of these are the interests which sift and evaluate the 'elements' of the schemes.

> In all these 'schemes' there is operating that kind of unwitting analysis which gives weight to certain elements of the whole....Most often of all, in human adult reaction it is a matter of the operation of persistent 'interests' (p. 302).

Here again, while attempting to think in Ward's terms of activity, the contradictory concept of memory *content* creeps in.

Bartlett's basic psychology is essentially Wardian; and his theory of remembering with its emphasis on interests is in harmony with this psychology. It is paradoxical that while this psychology is assumed by Bartlett, the use of schema which is consistent with it—the schema being a plan of the activity rather than a form into which content is put, or a force which causes the activity—is not developed in his theory. Instead, he attempts to substitute in his Wardian psychology the idea of schema as a mass of content which the interests serve to organize. The failure to develop the meaning of schema which is consistent with the psychology he develops, and the attempt to make the term harmonious with concepts in other fields such as these of Head and Rivers, is the ultimate basis of the constant confusion.

(*c*) *W. H. R. Rivers.* Bartlett's familiarity with Rivers's work and his own early and continuing interest in anthropology (4) add more confusion to the 'schema'. From his interest in the persistent group patterns and conventions which mould new activity into their own forms and change these forms very slowly, Bartlett works by analogy to the idea of the function of schema in individual activity.[1]

[1] That the analogy develops from Bartlett's interest in social processes and is taken over to the individual field is apparent from a study of Bartlett's social psychology—see § 3 of *Remembering*.

> The direction of a group's preferred, persistent tendencies constitutes a kind of lasting social 'schema' on the basis of which much successful constructive work in recall may take place (p. 264).
>
> It may be that social conventions, institutions, and traditions formed by persistent group tendencies constitute 'group schemata'; just as the individual images, ideas and trains of thought formed by persistent interests constitute 'individual schemata' (p. 299).

The schemata here have the meaning which Bartlett retains when he uses the concept to mean the forms which are imposed on new material and the forces which determine what the remembering activity will be.

A second systematic difficulty arises from Bartlett's relationship with sociology. According to his theory of remembering, interests, appetites, instincts and ideals influence remembering as they function in the schemata, but the social background, he considers, influences the individual directly and not as part of any psychological process. Remembering is determined, then, in two ways; first by the schemata, and secondly, and separately, by the social group. Therefore, the schema has to be given a meaning which will harmonize its effect with the effect of the social group. This is one source of the confusion which results when the schemata are thought of as forces.

(*d*) *Bartlett's early studies.* One other influence on the development of the concept should be noted—namely, Bartlett's own early experiments in perception. In an article, published in 1916, on "Some problems of perceiving and imaging"(5), he states that in perception the subject always makes an 'effort after meaning'. This meaning "depended greatly upon the prior experience of a subject and his constant interests or those of the moment". In his book of 1932, in reporting on these early experiments, he says, "I shall call this fundamental process of connecting a given pattern with some setting or *scheme* 'effort after meaning'" (p. 20). In this early notion of the influence of both past experience and interests in forming the scheme, the relationship of the concept with that of the 'apperceptive mass' is seen on one hand, and on the other its relationship with Ward's notion of continuity based on interests.

(*e*) *Summary.* In the above discussion the various influences which affected Bartlett's concept 'schema' have been indicated. These may be summarized briefly here. Although Bartlett's psychology is essentially that of Ward, he does not develop the idea belonging to this psychology of the schema as a 'plan of the activity'. He attempts rather to reconcile his use of the term with physiology on the one hand and anthropology on the other, and is thus led into employing it in several senses. The

schema is at times thought of as comparable to the forces exerted by social custom on behaviour, at times as a form which is like that both of a biological standard and of a social pattern, and at times as the content of memory and the apperceptive mass into which new material is received, even as new physiological impulses are registered against the stored effects of previous ones. The confusion which results makes systematic appraisal difficult and experimental development vague. The term itself must be clarified and defined. This is here attempted by further investigation of remembering as an active process; and through this will be determined where and how (if at all), the term best applies.

Accordingly, in the next section we turn to experimental studies of remembering. In these we shall consider at what point the concept of schema is necessary and in what way it may best be defined. Possibly any one of Bartlett's four ways of using the term might be developed, or perhaps some other use of it may be more adequate. The definition we select will be the one which develops from the experimental results.

IV. PURPOSE AND GENERAL METHOD OF THE EXPERIMENTAL INVESTIGATION

The purpose of these experiments is to continue the study of active remembering, and from the results give an interpretation of the notion of 'schema'.

In contrast with Bartlett's experiments with individual adults, these studies have all been made in classroom groups of school-children. The emphasis Bartlett placed on the use of 'real' material, that is, material with which the subject would most likely 'have to deal', has been maintained. This emphasis on 'reality' has been extended in the final experiment to include the use of the actual situation and usual method of learning the material, in so far as this could be done and a satisfactory degree of 'control' secured.

V. PRELIMINARY EXPERIMENTS

(i) *The influence of the 'difficulty' of the material on remembering*[1]

While this particular experiment arose originally as a by-product from another study, the results are recorded here to show a preliminary attempt to develop a definition of 'schema'.

Forty pupils in the fifth grade learned one poem a week for four weeks.

[1] Full reports of this and the following study have been published in the *British Journal of Psychology* (12, 13).

The learning periods were equal and the poems were all of the same length and form, but differed in the difficulty of their thought content. Four weeks after the last of these had been learned the children were asked to write "what they remembered of each poem, or what story it told". The number of words which were reproduced immediately after a learning period was taken as the objective criterion of the degree of difficulty the poems had for the subjects. With the difference in difficulty a second difference demonstrated itself. This may be considered a difference of "what the subject has made out of the material". With the simplest poem, the subject constructed a recall which, with a few minor changes, was closely determined by the given material. As the material became increasingly difficult, the subject's constructions showed more importation, less close adherence to the original and more individuality.

What place can we give to the notion of schema in this interpretation? There are two differences, one of content achieved, the second, of what the children have made from the material. If this second difference were called a schematic one, then the schema would have the meaning of "what the subject makes from given material". In the easy poems the schemata are closely related to the material; as the poems become more difficult the schemata show more invention and substitution. The use of the word schema to indicate this non-quantitative difference in what "the subject makes of material" is valuable in the interpretation of this experiment. The following experiments investigate whether this use can be maintained, further clarified and extended.

(ii) *The influence of age and social group on children's remembering*

Bartlett stresses the fact that social background influences remembering. In giving the same story to subjects from English and Indian cultures he found certain differences in recall.

> The variations (between the English and Indian subjects) seem to be in accord with group differences. Both the character and operation of group differences could be studied by the careful use of the Method of Serial Reproduction (p. 146).

In this experiment recalls were obtained from groups with a much smaller difference in their social background. The subjects were children from three schools in different districts of Toronto, with as widely different social backgrounds as possible in one community. Three stories, each approximately 250 words in length, but varying in type, style, and difficulty were used. The chain or serial reproduction method (p. 119) with 2–8 chains in each group for each story was used. The subjects read the story through once and immediately wrote what they remembered

of it. They then left the room and the subjects for the next link came in and each read one of the former subject's reproductions. A week elapsed between each story. The chain was carried on until either the subjects or the material were all used.

Differences in remembering among the groups were as follows:

(1) In the reports of the older children (14–15 years) more of the given material was retained than in those of the younger children (10 years). The younger ones imported and substituted their own ideas.

(2) The less stable social group gave many more modifications and importations than the comparable age group in the more stabilized social setting.

(3) Items having a high positive value of social meaningfulness were remembered most readily.

These results support the rather obvious claim that social group and age influence remembering. Have they any contribution to make to the clarification of the concept of schema? In considering this we must refer again to our section on the historical development of the notion. There it was pointed out that Bartlett considered remembering to be determined by the schema and, as a result of Rivers's influence, by the social group. This, we mentioned, led to some confusion; for remembering was determined first by a force which was psychological in nature, and secondly, by a force belonging to an anthropological frame of reference, namely the social group. From these results it is suggested this confusion may be clarified in this way. As in our first experiment we are concerned primarily not with quantitative differences, but with differences in "what the subject makes out of the material". These, we suggested, might be called schematic differences. Using the term in this way we need not think of schema and the social group as two separate forces influencing remembering, but rather that social background and age are factors which are reflected in the schemata or "what the subject constructs from the material". The schemata reflect the characteristics of the group as they are incorporated by the individual. The members of the heterogeneous group show more diversity in their schemata than do those of the stabilized group, the younger children show more originality than the older ones. And finally, using the term schema in this sense conveys the significance that age and social group do not influence the quantity of material remembered, but the form the activity takes.

Summarizing the results of these experiments in terms of the use we have assigned to the word 'schema', it may be stated that "difficulty of material, age and social group are all factors which influence the memory schemata".

A NEW APPROACH TO THE TESTING OF MUSICAL ABILITY

By BORIS SEMEONOFF

From the George Combe Psychological Laboratory, University of Edinburgh

I. *Introductory* (pp. 326–327).

II. *Preliminary experiments* (pp. 327–329).

III. *Experimental method and results* (pp. 329–337).

(1) *Basic assumptions* (p. 329).

(2) *General method of testing employed* (pp. 329–330).

(3) *A measure of musical knowledge: revised Kwalwasser test* (pp. 330–332).

(4) *Test material and results: Series A* (pp. 332–333).

(5) *Test material and results: Series B* (pp. 334–335).

(6) *General results from both series and from the revised Kwalwasser test* (pp. 335–336).

(7) *Conclusions from the correlations obtained* (pp. 336–337).

IV. *Résumé and discussion* (pp. 337–339).

(1) *Validity of the method of testing* (p. 337–338).

(2) *Possibility of improvement of the test material* (pp. 338–339).

(3) *Desirability of a further revision of the Kwalwasser test* (p. 339).

(4) *Applicability to other groups* (p. 339).

V. *Summary* (pp. 339–340).

References (p. 340).

I. Introductory

Tests of musical ability may be classified into two main types—those that measure proficiency resulting from education in music, and those that measure some innate capacity for the performance or appreciation of music. Existing test schemes have not always distinguished between these two types, which correspond to definitely different aims. Even some of those purporting to be of the second type have been cast in a form which has demanded knowledge on the part of the testee of musical terminology, or has otherwise placed individuals with musical experience at a distinct advantage. The purpose of the present paper is to suggest the possibility of the use of a new principle in the testing of one aspect of untrained musical ability.

In point of fact, the subjects of the present experiment all had some knowledge of music, however slight, and the material was chosen with this particular group in mind. The group consisted mainly of students of the Ordinary Psychology Class of the University of Edinburgh, but a few other people were included, all with a similar educational background. In addition, use was made of a small group of students in the Faculty of Music, who acted mainly as a control. It is necessary, therefore, to stress the fact that this paper describes an *approach* to the testing of musical ability. It was hoped that if the method proved successful with the student group, it might also be applicable to school children, down to quite an early age.

II. PRELIMINARY EXPERIMENTS

The method of testing used evolved gradually. Some hints were obtained from Schoen (3). Since many of the researches described in that book seemed to be lacking in general conclusions, two forms of questionnaire were drawn up, and tried out in consecutive terms with the Ordinary Psychology Class during session 1937–8.

In each case a series of gramophone records of music of various types was played, and the hearers reported their replies to certain questions. A good portable gramophone was used, and, except in a few cases, non-metallic needles were preferred to steel. It was thought that the variety of music available on gramophone records, and the assurance of constancy of performance, more than compensated for the lack of 'human element', since the writer's personal opinion was that such 'human element' might be prejudicial to one's judgement.

The form of questionnaire first used was as follows:

Name..... Date.....
Do you sing, or play any musical instrument?.....
Do you attend concerts? If so, of what kind?.....
After each selection is played, answer the following questions: (*a*) Did you like it? (*b*) Can you identify it (guess if you like)? (*c*) What impression do you think the composer is trying to convey? (*d*) Did the music give you any emotional experience?

Spaces for replies followed.

An examination of the results led to the following conclusions:

(i) A large proportion of the listeners reported that a piece of music was vaguely familiar, but that they "couldn't place it", or "could never remember the name of a tune". This suggested that one could be reasonably sure that even not totally unfamiliar music would be identified by only a small number of hearers.

(ii) The variety of 'impressions' reported in answer to question (*b*) suggested that this diversity might be used either as a basis of classification, or as an indication of understanding on the part of the listener.

(iii) It appeared that the emotional element, at least under test circumstances, was considerably less important than had been expected, and normally did not prevent the subject from responding in the other ways required.

The experiment was repeated three months later, using different records, and a modified form of questionnaire, as follows:

Name..... Date.....
Do you sing, or play any musical instrument?.....
Are you fond of music?.....
What kinds do you prefer? .....
After each selection is played, answer the following questions: (*a*) Did you like it? (*b*) Have you heard it before? (*c*) If so, can you identify it? If you wish to guess, do so. (*d*) What did you think about while the music was being played?

Spaces for replies followed, as before.

Responses to the second form of questionnaire confirmed the first of the conclusions from the previous experiment. Analysis of the replies to question (*d*) of this series further indicated that most listeners spontaneously read a meaning into the music they heard. These 'meanings' for the most part took the form of 'personal' (i.e. egocentric) or what might be called 'general' associations. These latter were considerably the more frequent, especially among men. In all, 420 'general' association responses were recorded, as against 259 'personal' associations. These figures do not include cases of imagery, impulse, emotional experience, like or dislike, or remarks concerning the style or execution of the music, all of which occurred much less frequently. The element of 'meaning' of a generalized character thus seemed to play a considerable part in the experience of the listeners. Finally, a cursory examination of the general questions at the head of each questionnaire suggested that the ability to recognize correctly the composer's intention or other implications of a musical composition depended on the hearer's musical ability, or at least his musical 'background'. The general questions referred to of course explored only the degree to which the subject was *in contact with* music, and gave no indication of actual ability. Other things being equal, it would appear that the musically gifted person would also be interested in and acquainted with music, but of course other things are seldom equal, and it may be that many people with a potential ability to produce or appreciate music are prevented from doing so as a result

of lack of opportunity or unsuitable musical education. This view was borne out in the present case by examining the results of an experiment similar to those reported, performed with a small group of students of an Adult Education class in Psychology.

III. EXPERIMENTAL METHOD AND RESULTS

(1) *Basic assumptions*

The theoretical basis of the main study, to be discussed presently, is therefore as follows:

Musical ability implies something more than simply the ability to sing or play an instrument, or to discriminate between simple or complex sensory impressions such as form the raw material of music, and the basis of certain existing tests. It is a widely accepted fact that what distinguishes the really first-rate musician from the ordinary competent orchestral performer is not so much superior technique as interpretative power; similarly, a person with next to no ability as an executant may be a highly gifted critic or even conductor. (In addition it is significant that some of the greatest composers have been indifferent executants.) Interpretative power, then, may be regarded as an important part of musical ability, and this, to the writer's knowledge, has never been separately tested, at any rate on systematic lines.[1]

The question next arises how far one is justified in expecting uniformity of interpretation, and so to what extent one can apply the predicates 'right' and 'wrong' to an interpretation. This, however, is a difficulty which need not invalidate the method of testing employed, although it does limit to some extent the choice of material.

(2) *General method of testing employed*

The method used was to ask the hearer to select from a group of alternatives that interpretation which he thought best fitted the music, this limitation of response being a principle widely used in intelligence testing and elsewhere. Although it might appear at first sight that only music with a definite 'programme' could be used, this is not necessarily the case. If one accepts the broadest implications of Howes's contention(1) that "all music of any value is at bottom programme music, that is, it is a commentary on life, and not something divorced from life", then any musical work in respect of which the composer's intention is

[1] Since writing the above, however, I have heard papers, at present unpublished, describing test-batteries including tests of interpretative ability, using methods similar in some respects to that of the present experiment.

definitely known may be used as test material. It may, of course, be objected that the composer may have failed to give adequate expression to his intention, but this is hardly a valid objection if one limits oneself to music on which the consensus of competent opinion shows complete agreement. An experimental check on this point is described later (§ IV (1)).

The musical material used was selected in accordance with these assumptions, the choice being limited to such gramophone records as were readily accessible to the writer. Ten records was found to be a convenient number to constitute a series, this being the same number as was used in the preliminary experiments. It was possible to play these, allowing for explanation, pauses, etc., in just under an hour.

(3) *A measure of musical knowledge: revised Kwalwasser test*

It was, of course, necessary to use some sort of criterion of musical ability in evaluating the results obtained. Unfortunately, it was not possible to make use of the Seashore Measures of Musical Talent(4) or any other recognized tests, but in any case considerable doubt was entertained regarding their validity for the present purpose. It was possible, however, to obtain an indication of the 'musical sophistication' of the subjects by means of a questionnaire. A fairly suitable test was available in the Kwalwasser Test of Music Information and Appreciation(2), but this was rather long for the time at the writer's disposal, and it also seemed possible to improve upon it in other respects. Accordingly, a radically revised abridgement of the Kwalwasser test was made. The relation between this revised form (RK) and the original (Kw) may be seen in Table I. The last column of figures in the table gives the number of questions not contained in the original form of the test.

The general plan of the revision was (*a*) to omit as far as possible questions of a more intimately personal character, e.g. "Clara Wieck became the wife of Schumann", and those in which 'general knowledge' might be said to play a large part, e.g. proper names of which the form revealed the nationality; (*b*) to include a few questions dealing with vocal music, and modern and light music, not excluding jazz. The Kwalwasser test ignores these almost entirely.

Kwalwasser's Test 1 was omitted because many of the names of performers were unlikely to be familiar to European subjects, or had ceased to be prominent. Test 3 requires uncontrolled naming of composers, which was considered to be contrary to the general plan of the test method

used. Test IV of the revised form, devised by the present writer as a substitute for Test 3, was constructed on the multiple-answer plan, and tested detailed knowledge of well-known compositions. In additions to these alterations, minor changes were made in the directions for Tests 2, 5, and 6, and in the examples given.

A correlation of $0{\cdot}83 \pm 0{\cdot}019$ was obtained between the 'retained' and the 'new' items.

The test booklet was provided with a title-page (referred to herein as 'page 1'), which was designed to measure actual contact with music by means of replies to a series of questions. Its purpose was to detect

Table I. *Relation between original and revised forms of Kwalwasser Test of Music Information and Appreciation*

Test no.			No. of items		
Kw	RK	Title of test	Kw	RK	New
1		Classification of artists	20	—	—
2	I	Nationality of composers	20	10	4
3		Composers of famous compositions	30	—	—
4	II	Classification of composers	10	5	2
5	III	General knowledge of composers and compositions	50	15	7
	IV	—	—	12	12
6	V	Production of tone on orchestral instruments	10	5	0
7	VI	Classification of orchestral instruments	10	6	1
8	VII	General knowledge of instrumentation	50	12	2
9	VIII	General knowledge of musical structure and form	50	15	5
		Total	250	80	33

those individuals whose interest in music was of a more active and possibly non-technical kind, but naturally it was difficult to interpret these replies objectively. Thus, for example, in reply to a question "Do you play any musical instrument?" those who say they play 'a little' frequently play considerably better than many who reply simply 'yes'. Correlations between a crude numerical measure based on 'page 1' and various other variables are noted later (§ III (7)).

Both 'page 1' and the actual results of the revised Kwalwasser test are thus a little difficult to interpret. Any conclusions drawn depend to some degree on considerations such as to what extent potential musical ability overcomes obstacles put in the way of its development, and, conversely, to what extent training in or theoretical knowledge of music may compensate for lack of original ability. However, acting on the assumption that an individual who had gone far enough to undergo

professional training in music would be almost certain to possess genuine ability as well as interest and knowledge, the writer decided to ask permission from the Edinburgh University Faculty of Music to try out the tests with students of the Interpretation Class. A group of twenty to thirty students was made available, and although the number was small, it acted as a useful control group.

(4) *Test material and results: Series A*

Originally it was not intended to draw up more than one series of test records. Accordingly the set referred to as Series A consisted of a first selection of records considered by the writer to be the most suitable of those available, at the same time giving sufficient variety. Series A comprised the following records:

(1) Minuet from *Don Giovanni* (Mozart): harpsichord with orchestral accompaniment.
(2) "By the Waters of Minnetonka": trio—piano, violin and cello.
(3) "Volk ans Gewehr" (modern German song): male chorus with military band accompaniment.
(4) "The Elephant" from *The Carnival of Animals* (Saint-Saëns).
(5) "Finlandia" (Sibelius)—first half.
(6) "Kolj Slavenj" (Bortniansky): Russian church music—Don Cossack Choir.
(7) "Pastorale" (Stravinsky): violin and wind instruments.
(8) Trio in B♭, Op. 99 (Schubert)—1st movement, closing portion.
(9) "La Fille aux Cheveux de Lin" (Debussy): violin arrangement.
(10) "I have attained the power" from *Boris Godunov* (Moussorgsky).

Where no setting is mentioned, it is to be understood that the original form of the composition was used. In the cases of the vocal selections, it was found that in no case were the hearers able to deduce the required answer from an understanding of the words.

The report sheets issued were headed as follows:

Name..... Date.....

After each record has been played, answer the following questions: (*a*) Did you like it? (*b*) On what factors did your answer to (*a*) depend? (*c*) Underline the word or phrase completing the description of the music given below.

Although use was also made of the replies to (*a*) and (*b*), the chief interest lay in the replies to question (*c*), samples of which are as follows:

(1) A dance measure, suggestive of (solemnity dignity gaiety abandon).

(10) A soliloquy: the singer's frame of mind is one of (grief despair dissatisfaction defiance).

As will be seen from Table II, these examples represent fairly well the extremes of difficulty involved.

Table II. *Series A: Percentage frequency of responses recorded, of more than one response, and of omissions*

No.	Responses %								Over 1	Omissions
1	solemnity	3	*dignity*	*88*	gaiety	6	abandon	1	4	1
2	moonlight	6	*rippling water*	*73*	bird song	8	trees	3	7	1
3	*determination*	*79*	triumph	10	exhilaration	7	solemnity	—	1	2
4	tiger	15	camel	22	*elephant*	*46*	tortoise	13	—	4
5	sub-tropical	6	temperate and fertile	1	*wild and rugged*	*81*	industrialized	11	1	—
6	resignation	52	thanksgiving	13	communion	24	*praise*	*6*	3	2
7	*pastoral scene*	*28*	serenade	2	eastern bazaar	41	exotic dance	26	1	1
8	austere	2	headstrong	14	carefree	32	*romantic*	*44*	7	3
9	handsome youth	1	*beautiful young girl*	*32*	mature woman	5	gracious old lady	58	—	4
10	grief	27	despair	35	*dissatisfaction*	*15*	defiance	13	6	4

Of the records in Series A, Nos. 1, 2, 3, 6 and 7 had been tried out in the preliminary experiments. Table II shows for each item the percentage of times each of the given responses was recorded, together with, in the last two columns, the percentage of cases in which more than one answer, and no answer, respectively, were given. Correct replies are italicized. The various alternatives should give a fair indication of what each question was. It will be noticed that all possible answers were received, with one exception. These percentages are based on 142 cases—119 members of the Unselected group and 23 of the Musical group. In some cases rather interesting differences between the groups were obtained. These are shown in the first two columns of Table III; the two later columns show similar data for percentages *liking* each selection.

Table III. *Series A: Percentage of Unselected group and of Musical group correct in each selection and liking each selection*

Record no.	Percentage correct		Percentage liking	
	Unselected	Musical	Unselected	Musical
1	95	96	92	78
2	75	58	82	17
3	84	50	61	48
4	37	92	47	96
5	80	87	83	82
6	6	4	65	14
7	25	35	55	64
8	36	58	70	100
9	22	83	68	100
10	19	0	55	50

(5) *Test material and results: Series B*

Since the number of test items was too small to obtain a split reliability, it was thought desirable to try a retest by means of a Series B. Two chief difficulties stood in the way of this: (i) as already indicated, the pick of the writer's records had been used for Series A, and (ii) difficulty was experienced in obtaining again the services of all the subjects who had listened to the first set of records.

The second of these difficulties was partly overcome by asking for volunteers, and eventually, when accidental circumstances had taken their toll, twenty-six subjects remained from the Unselected group, plus a slightly different group from the Interpretation Class. This of course left a rather small number for purposes of correlation, but even so, significant results were obtained, as will be seen. It should also be noted that the 'Unselected' group became rather less unselected, since presumably the students who volunteered were those who had more than the average interest in music.

The other difficulty was also overcome as well as circumstances permitted, and a Series B was drawn up in which each item was matched for type of music, if not for difficulty, with one in Series A. Series B comprised the following records:

(1) "Song of the plains" (modern Russian song): male chorus with band accompaniment.
(2) "The wind in the trees" (Briccialdi): flute.
(3) Polonaise from *Eugen Onegin* (Tchaikovsky).
(4) "Sérénade Espagnole" (Chaminade): violin.
(5) Symphony No. 1 in C maj. (Bizet)—4th movement, first part.
(6) "Tortoises" from *The Carnival of Animals.*
(7) "How beautiful those eyes" (Russian sentimental ballad): mezzo-soprano.
(8) "On hearing the first cuckoo in spring" (Delius)—first part.
(9) "Pimen's monologue" from *Boris Godunov.*
(10) "Poet and Peasant" Overture (Suppé).

Of these, only No. 2 had been tried out previously.

Tables IV and V show the data for Series B corresponding to those for Series A already given. Much less significance can be attached to these, in view of the very much smaller Unselected group which listened to Series B, and of the inferior musical quality of some of the records in this series. It will be noticed that the proportion of omissions is considerably higher; practically all of these came from the 'new' members of the Musical group.

Table IV. *Series B: Percentage frequency of responses recorded, of more than one response, and of omissions*

No.	Responses %								Over 1	Omissions
1	youth organization	13	infantry	4	*cavalry*	*71*	artillery	4	—	9
2	bee's flight	45	cloud of insects	36	rushing water	2	*wind*	*7*	—	9
3	gaiety	52	*stateliness*	*9*	boisterousness	28	grace	—	6	4
4	lullaby	6	*serenade*	*17*	aubade	50	pastoral scene	13	4	11
5	(age)									
	below 20	*31*	c. 30	44	40–45	11	over 60	2	4	7
6	crocodile	22	*tortoise*	*50*	bear	9	elephant	6	—	13
7	joy	2	passion	2	despair	19	*resignation*	*71*	—	7
8	*spring*	*33*	summer	32	autumn	22	winter	6	—	7
9	dejected	50	defiant	2	*old and weary*	*41*	self-confident	—	7	—
10	big city	2	mountainous	15	tropical	—	*rural*	*70*	2	11

Table V. *Series B: Percentages of Unselected group and of Musical group correct in and liking each selection*

Record No.	Percentage correct		Percentage liking	
	Unselected	Musical	Unselected	Musical
1	77	63	77	38
2	8	7	69	3
3	8	11	62	78
4	15	18	85	44
5	23	39	81	85
6	35	64	58	70
7	77	63	58	0
8	4	63	73	93
9	42	39	65	58
10	73	71	85	56

(6) *General results from both series and from the revised Kwalwasser test*

Highest, mean and lowest scores on all measures for the two groups are given in Table VI. The upper line of figures in each represents the Unselected group; the lower line, of italicized figures, the Musical group. Since the numbers tested and the differences obtained are both small, little absolute reliance can be placed on the data contained in this table. Nevertheless, some of the results are sufficiently striking and suggestive to form the basis of the conclusions given below.

The various correlations which were thought to be worth finding are given in Table VII. The figures in the second column are for the whole group, those in the first for the Unselected group only, since in some cases it seemed preferable to consider this group by itself. It will be noticed that some peculiar shifts take place when the Musical group is included. This is particularly so in the case of the correlations between A and B on the one hand, and 'D' and 'E' (see below) on the other. In

spite of the considerations noted above, it will be seen that these correlations satisfy the minimum requirements for statistical significance, though considered as reliability coefficients they are of course very poor. (At one stage of the investigation an attempt was made to redistribute the records in the two sets into a difficult and an easy series. The new correlations thus obtained are those between 'D' and 'E'.)

Table VI. *Ranges of scores: both groups (see text)*

	Highest	Mean	Lowest
Revised Kwalwasser	58	31·2	8
	76	*64·7*	*49*
Scores: Series A	9	4·8	2
	8	*6·2*	*4*
Scores: Series B	6	3·7	1
	8	*4·7*	*1*
Liking: Series A	10	6·1	3
	9	*6·3*	*3*
Liking: Series B	10	7·3	3
	8	*5·3*	*2*

Table VII. *Correlations*

	Unselected	All
Kwalwasser and 'page 1'	0·45 ±0·048	—
Kwalwasser and A	0·28 ±0·061	0·39 ±0·052
Kwalwasser and B	—	0·44 ±0·087
Kw and A *liking*	0·20 ±0·064	0·02 ±0·059
Kw and B *liking*	—	−0·30 ±0·098
'page 1' and A	−0·05 ±0·064	—
'page 1' and B	0·27 ±0·127	—
'page 1' and A *liking*	0·31 ±0·056	—
'page 1' and B *liking*	0·40 ±0·111	—
A with B	0·43 ±0·112	0·37 ±0·090
'D' with 'E'	0·52 ±0·099	0·32 ±0·093
A and A *liking*	0·16 ±0·061	0·15 ±0·056
B and B *liking*	0·27 ±0·123	−0·05 ±0·096
Liking: A and B	0·74 ±0·062	0·72 ±0·051

(7) *Conclusions from the correlations obtained*

Other main points which seem to emerge from the correlations are as follows:

(i) *Liking* seems to be an independent and fairly constant factor in this test situation. It was found difficult to devise a technique whereby it could be discovered whether 'understanding' of a composition was correlated with liking for it. Correlating the number of selections liked with the number correct did not really meet the case, and the distributions obtained made it almost impossible to calculate association coefficients for each record separately. Nevertheless, it was found that negative or very low association coefficients were found mainly in cases where either there was a widespread antipathy to music of that type

(e.g. A 3), or the music was definitely in an unfamiliar idiom (e.g. A 7, A 9). On the whole it seemed possible to draw the general conclusion that, other things being equal, that music was enjoyed which was correctly understood.

(ii) In some respects 'page 1' was a better or more useful measure of 'musical sophistication' than were the questionnaire scores. This is seen especially in the case of the correlations with *liking*. While there was here no marked connexion between musical knowledge (as measured by the questionnaire) and liking, experience of music of a more direct kind (as measured by 'page 1') correlated with it positively, to a degree statistically significant. On the whole, the indications are that knowledge of music widens the scope of what one can enjoy—a point contrary to the view that the untaught show lack of discrimination, while the sophisticated are unduly eclectic. It is noticeable, however, that in the case of Series B, there is a *negative* correlation with the Kwalwasser scores—as might have been expected in view of the low musical value of some of the B records.

IV. RÉSUMÉ AND DISCUSSION

The following discussion restates the main standpoint of the investigation, and amplifies certain points arising from the experimental procedure and the results.

(1) *Validity of the method of testing*

The object of the experiment was to test musical ability on the basis of the assumption that one way in which the musically gifted individual shows his superiority is in correctly identifying the composer's intention or other authenticated content of a musical composition, whether of a 'programme' character or otherwise, provided that that content is sufficiently clearly expressed. That such an ability exists, the results of the present experiment seem to show. How far it is correlated with other aspects of musical ability cannot be known until results from really satisfactory test material are available. Although many of the selections used were correctly interpreted by only a small proportion of the hearers, it was possible to check the correctness of what may be called the 'official' interpretation as follows:

From the records used, a selection of six was made on which low scores had been obtained; these, together with three others not previously heard, were played to sixteen members of the Unselected group who had listened to both series. Before playing each record, the writer gave some

information regarding its title and content, in some cases pointing out certain features which were thought to indicate why the 'official' interpretation was correct and the others wrong. Among other questions, which served partly as 'buffer' questions, the subjects were asked to state whether they thought the suggested interpretation was borne out by the music. This procedure of course lays itself open to the charge of suggestion, but from his personal knowledge of the individuals concerned, the writer does not consider that this factor had much influence on the results obtained. Table VIII shows the numbers of affirmative answers received against the numbers of correct interpretations previously recorded by these subjects. Allowing for a noticeable (and possibly conscious) obstinacy on the part of one or two individuals, and for the few 'reversals' recorded, the results may be taken as fairly reassuring.

Table VIII. *Number of correct responses on first hearing, and number of agreements and non-committal responses on rehearing: sixteen subjects*

Record		First hearing. Correct	Rehearing	
			Agreement	Non-committal
B6	Tortoises	4	9	—
A4	Elephant	6	7	—
B5	Symphony	3	14	1
A6	Kolj Slavenj	1	9	1
A7	Pastorale	5	9	—
B4	Sérénade	2	9	4
	Total	21	57	6

(2) *Possibility of improvement of the test material*

The material might be generally improved in the following way: Each proposed test item would be submitted to as large a body as practicable of really competent critics, and any item on which insufficient unanimity was forthcoming would be discarded. Thus, for example, in the case of the present Series A, a member of the Faculty of Music refused to accept the applicability of the question asked to the Schubert Trio in B♭. On the other hand, a large proportion of the Musical group gave the expected response. How far this was due to actual knowledge of the work would be hard to determine, but it serves to re-emphasize the desirability of unfamiliar material in a test situation such as this. When another large group of subjects becomes available it is hoped to repeat the experiment with practically the same material. If some of the alternatives suggested were altered, possibly only one record (No. 6) would require to be discarded from Series A. Nos. 1 and 2 could be left as shock absorbers, and No. 3 could be made more discriminative by

substituting a more tempting alternative for 'solemnity'. Further adjustments or replacements could also be considered.

(3) *Desirability of a further revision of the Kwalwasser test*

In any investigation of musical ability, at any rate in the case of adults and older children, the use of a check of musical knowledge, such as the Kwalwasser test, seems essential. Experience in the present instance indicates that it would be desirable to get rid, as far as possible, of the 'Right minus Wrong' method of scoring, which is used in three sub-tests in both the original Kwalwasser form and the writer's revision. Tests V and VI might be improved by making them a little more difficult and by confining Test V to instruments other than those of the symphony orchestra. The 'page 1' questions might possibly also be made more systematic.

(4) *Applicability to other groups*

In conclusion the writer wishes to stress the fact that the experiment reported is to be regarded as an *approach* to a method of testing, and, further, as an approach from what might be called 'the top'. Ultimately, it might be possible to adapt the method to the testing of quite young children, and the writer hopes, if possible, to work down gradually through the various age groups. Interesting results might also be obtained by testing groups of adults with comparatively little experience of music, though in this radio age such people, whether adults or children, would no doubt be hard to find. Quite apart from the diagnostic value of the test method, the writer has more than once heard the opinion expressed that it could be of considerable use in the teaching of Musical Appreciation.

V. SUMMARY

1. The experiment described was designed to meet the need for a test of musical ability of an untrained type, as distinct from proficiency resulting from education in music.

2. Preliminary experiments suggested that interpretative processes are an important aspect of the experience of, and enjoyment derived from, listening to music. Different individuals were found to possess different degrees of interpretative power; it was therefore thought possible to test this aspect of musical ability quantitatively.

3. Two series, each consisting of ten gramophone records of music of varied types, were drawn up, and report-sheets prepared. The listener was required to state whether he liked each record, and to select the

correct interpretation from a group of four alternatives. The subjects were students of the Faculties of Arts and Music. Despite deficiencies in the material, statistical treatment of the results indicated that ability to interpret music was fairly constant for the individual, and relatively independent of musical training and experience.

4. A check on the legitimacy of postulating a correct interpretation was obtained. It is now hoped to extend the test method to other age groups. Suggestions for the improvement of the test-material are proposed.

I wish to express my indebtedness and thanks to the following:

To the publishers of the Kwalwasser test for permission to publish results based on my revision of that test; to Prof. Sir Donald Tovey and Dr Mary Grierson for permission to test in the Faculty of Music; to all those who volunteered their services as subjects, or gave other assistance.

REFERENCES

(1) Howes, F. (1926). *The Borderland of Music and Psychology.* London.
(2) Kwalwasser, J. (1927). *Kwalwasser Test of Music Information and Appreciation.* State University of Iowa, Extension Division.
(3) Schoen, M. (ed.) (1927). *The Effects of Music.* London.
(4) Seashore, C. E. *Measures of Musical Talent.* Columbia Phonograph Co.

(*Manuscript received* 18 *May* 1939)

THE NATURE OF THE PRACTICAL FACTOR (*F*)

By E. J. J. PRICE

(*From the Oxford University Institute of Experimental Psychology*)

I. *Introduction* (pp. 341–342).
II. *Method of experiment* (pp. 342–346).
(1) *The test variables* (pp. 342–346).
(2) *Subjects* (p. 346).
III. *Results* (pp. 346–349).
(1) *Averages and standard deviations* (p. 346).
(2) *Correlations and factor analyses* (pp. 346–349).
(3) *Results of a multiple-factor analysis* (p. 349).
IV. *Conclusion* (pp. 349–351).
References (p. 351).

I. Introduction

Two recent studies have made some inroads into the common-factor structure of intelligence tests of a practical and of a primarily perceptual kind respectively. Alexander (1935) has demonstrated that performance tests involve a factor over and above the Spearman *g*-factor in these tests; he called it '*F*' and proceeded to argue for a concept of 'practical intelligence' in terms of his findings. Koussi (1935) and later Smith (1937), have likewise isolated a factor over and above *g*, called *K*, in pencil-and-paper tests which require the perception of geometric and similar two-dimensional shapes, and the eduction of space relations.

Now there are obvious similarities between the two kinds of tests supplying these factors *F* and *K* respectively, and one wonders at once whether the two factors are not, after all, one and the same. Performance tests, and paper-and-pencil 'form-relations' tests, both require shapes and their parts to be cognized, and space relations to be educed about them. The same kind of 'mental set' is involved too. One of the best *K*-factor tests, that called 'Fitting shapes' (Brown & Stephenson (1933)), is merely a set of form-boards represented on paper. The present study seeks to determine, therefore, whether or not *K* and *F* factors are one and the same. The problem is as follows: if a matrix is constructed which contains a number of performance tests, a number of pencil-and-paper 'space' or 'form-relations' tests, and a number of reference *g*-tests, will

two factors (other than specifics), Spearman *g* and one other, be sufficient to account for the correlations? The following evidence suggests that the answer is in the affirmative.

II. Method of experiment

(1) *The test variables*

The following tests were used for the matrix required:

Reference g-tests:

No. 1. Stephenson verbal intelligence test. Book I.[1]
No. 2. Stephenson non-verbal intelligence test. Book II.
No. 3. Stephenson non-verbal intelligence test. Book III.

Performance tests:

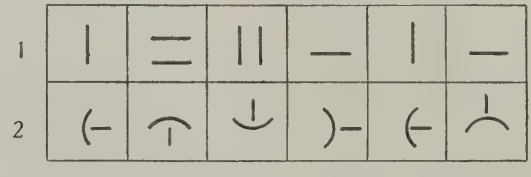
No. 4. Modified Dearborn-Anderson (see references below).
No. 5. Cylinder construction.
No. 6. Kohs' block design (1920).
No. 7. Woolley blindfold (see references).
No. 8. Alexander's 'Passalong' (1933).

Space-factor pencil-and-paper tests:

No. 9. Fitting shapes.
No. 10. Shape completion.
No. 11. Pool image.

The verbal intelligence test consists of several pages of verbal analogy test units, in four different forms; since it is the only verbal test in the matrix it can be used here as a 'reference' test for Spearman *g*-factor. The non-verbal intelligence tests are of the kind used by Koussi (1935), and in the Spearman-Holzinger study (1937), and have been shown not to involve *K*-factor. They will be described elsewhere, and it is sufficient to give a sample only of each at this juncture. Test No. 2 contained seven pages of problems of the kinds shown below:

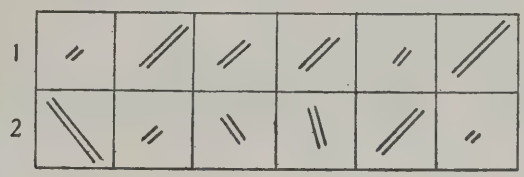
Pages 1 and 2: non-verbal classification test units.
(1) Select four of the six drawings that are most obviously alike in some way:



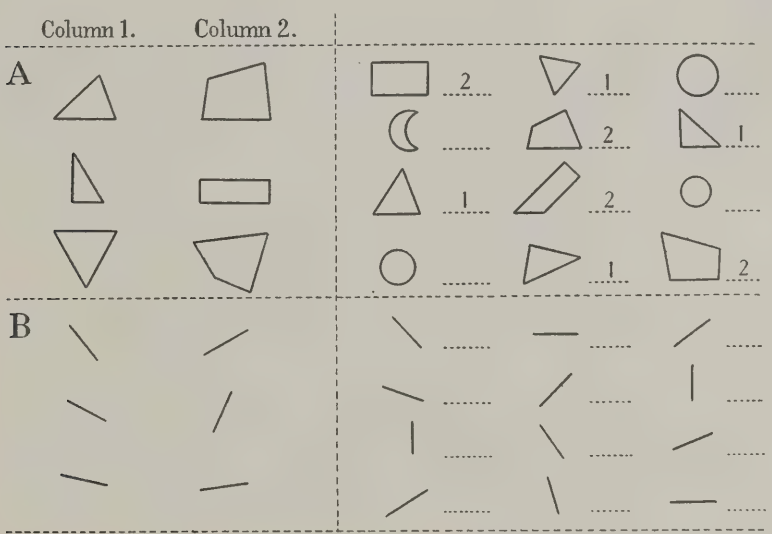
[1] Tests Nos. 1, 2, 3 are standard tests, used in all factor work in the Institute.

Pages 3 and 4: non-verbal series test units.

(2) Two of the six drawings have been interchanged; when put in their correct place the six make a regular series with respect to some feature of the drawings:



Pages 5, 6, and 7. Classification. The three drawings under A are alike in some way; those under B are alike in a different way; find four drawings at the right that are like those under A, and four like those under B:



Test No. 4 contained eighteen items, similar to those used in the Dearborn-Anderson form-board test. Each item consists of a small board (12 × 12 in. approx.) with an inset, into which three pieces of wood can fit; these three pieces serve in turn for all eighteen boards. The subject was given 15 sec. in which to attempt each board, and the number of pieces correctly inserted in that time was taken as the score for that board.

The Cylinder Construction test was devised for this study, with the object of seeing how far *F*-factor extended into three-dimensional constructions. A number of pieces are given which, when put together in a rack ready for the purpose (see Fig. 1), make up a solid wooden cylinder, 6 in. high, 2 in. diameter. Six cylinders have to be con-

structed in all; the number of pieces correctly assembled in each in one minute is taken as the score for that item.

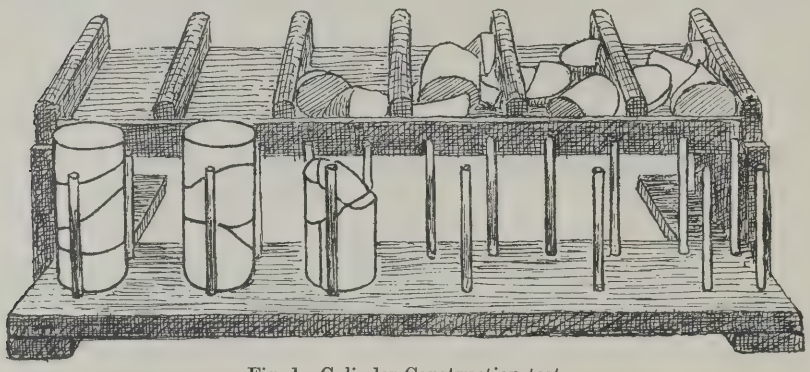


Fig. 1. Cylinder Construction test

Kohs' Block Design Test was the standard test, used by Alexander to demonstrate *F*-factor. Alexander's Passalong test was that described in the *British Journal of Psychology* (1933), except that in making it we decided to make the pieces larger, for ease in manipulation. The Woolley Blindfold test was also made in the laboratory. It consists of twelve wooden shapes fixed on a board (see Fig. 2). Each of these shapes

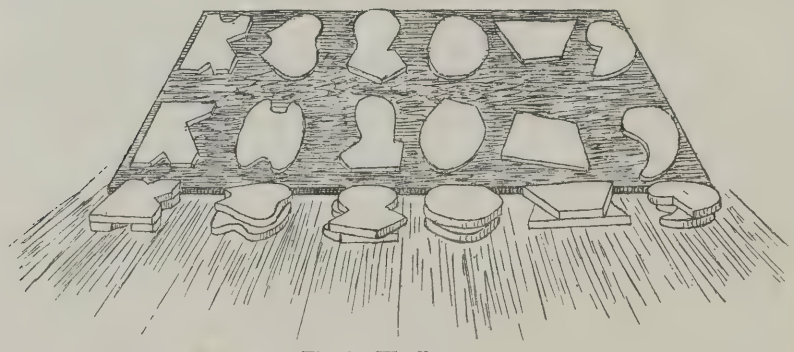


Fig. 2. Woolley test

has an unattached replica. The subject saw neither the fixed pieces nor their counterparts. Both written and verbal explanations of what he was asked to do were given him, and when he fully understood them, he was blindfolded. The test board was then placed before him, and he was allowed to pass his hands over the pieces fixed on the board for 20 sec. At the end of that time he removed his hands from the board, and was given one of the loose pieces to handle for 10 sec.; he was forbidden to touch the board while handling the loose piece. At the end of the 10 sec. the piece was taken away from him. He was then asked to identify its counterpart on the board; for this he was allowed another 10 sec. He was of course blindfolded the whole time, and had

to solve the tests by touch, i.e. by haptic sense. Two marks were given for each piece correctly identified. It will be seen that there is a certain resemblance between the upper and lower members of each pair of pieces on the board: if a subject was successful to the extent of finding the right pair, but the wrong member of it, he was awarded one mark.

Samples of the Fitting Shapes test, the Shape Completion tests, and the Pool Image test are given in Figs. 3, 4 and 5.

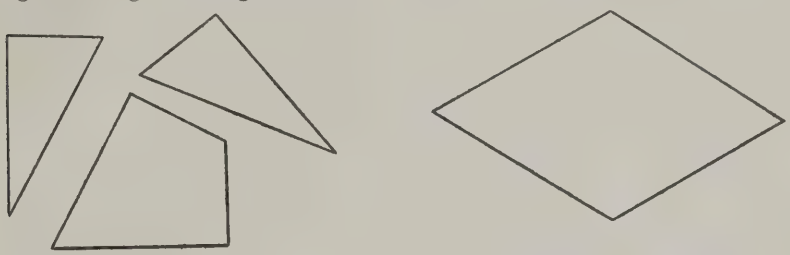
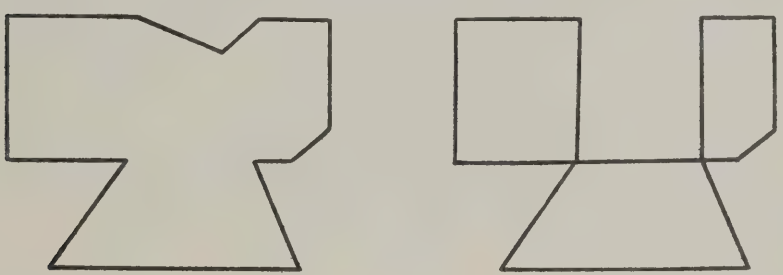


Fig. 3. Fitting Shapes test. Draw lines to show how the right-hand figure should be cut to give the shapes on the left



Standard Shape

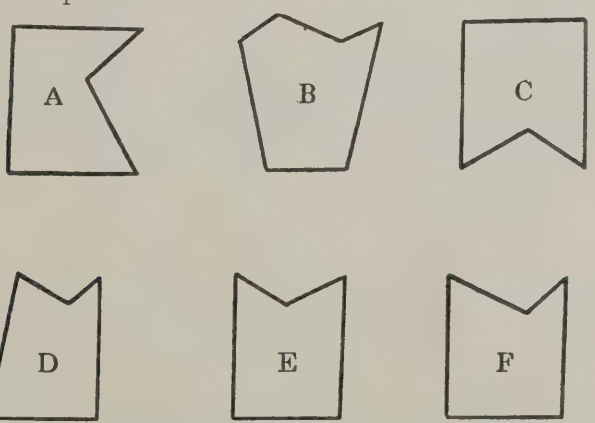


Fig. 4. Shape Completion test. Item 1. Select *one* of these pieces, which, fitted to the right-hand shape above, makes the *standard shape*

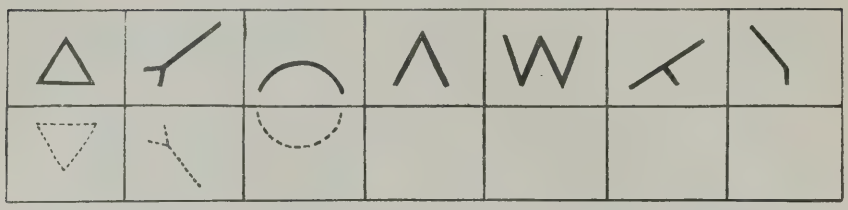


Fig. 5. Pool Image test. Imagine the blank squares to represent pools of still water, and draw in each what you think would be the reflection of the figure above it

(2) *Subjects*

Eighty-five University men students were tested individually by the author. All were between 19 and 22 years of age, were selected at random, and attended either Oxford University (60) or University College, London (25). None had done performance tests of this kind before. A few, of course, had sat for verbal intelligence tests at one time or another. The number of subjects, whilst not large, is sufficient for the present purpose, and represents a considerable amount of individual testing.

III. Results

(1) *Averages and standard deviations*

The mean scores and standard deviations for each of these tests, for the population of eighty-five students, are shown in Table I.

Table I

Test	Mean	Standard deviation
1. G–1	61·88	13·50
2. G–2	66·24	13·35
3. G–3	13·65	3·52
4. Dearborn-Anderson	37·38	9·08
5. Cylinder Construction	16·25	4·88
6. Kohs' Block Design	18·60	5·62
7. Woolley	16·08	3·45
8. Passalong	10·70	4·86
9. Fitting Shapes	35·16	13·27
10. Shape Completion	10·60	4·15
11. Pool Image	15·09	4·03

(2) *Correlations and factor analyses*

Table II gives the product-moment correlations for the eleven variables. They are positive, except for one which is insignificantly negative.

The simplest factor study for our purpose depends upon accepting the *g*-reference values (Tests 1, 2, 3) as the basis of an analysis. This amounts in effect to a Bi-factor analysis, such as Holzinger advocates (1937), which has in fact been the basis of much Spearman analysis. The

three g-tests are used as 'reference values', and their own g-saturations are determined in terms of each other;[1] the g-saturations of the other tests are then calculated one at a time in terms of these three as reference values.[2] These two steps give the factor saturations shown in column (i)

Table II. *Intercorrelations of eleven tests* ($N = 85$)

	1	2	3	4	5	6	7	8	9	10	11
1		0·493	0·399	0·106	−0·008	0·306	0·284	0·101	0·290	0·149	0·171
2			0·463	0·285	0·328	0·478	0·373	0·386	0·464	0·358	0·384
3				0·251	0·343	0·479	0·169	0·166	0·267	0·308	0·302
4					0·468	0·547	0·133	0·144	0·576	0·535	0·500
5						0·442	0·230	0·191	0·411	0·447	0·427
6							0·341	0·209	0·717	0·594	0·585
7								0·132	0·143	0·267	0·250
8									0·205	0·206	0·105
9										0·486	0·679
10											0·423
11											

1. G–1.
2. G–2.
3. G–3.
4. Dearborn-Anderson.
5. Cylinder Construction.
6. Kohs' Block Design.
7. Woolley.
8. Passalong.
9. Fitting Shapes.
10. Shape Completion.
11. Pool Image.

[1] The g-saturation of test 1 in terms of 2 and 3 is given by:

$$r_{1g} = \left\{\frac{r_{12} \cdot r_{13}}{r_{23}}\right\}^{\frac{1}{2}}.$$

Similarly for tests 2 and 3.

[2] The g-saturation of any of the tests 4–11 is calculated by adding them one at a time to the matrix for variables 1, 2, and 3. This is most conveniently done by using an adaptation of the Spearman formula (21) viz.

$$r_{ag} = \left\{\frac{A^2 - A'}{T}\right\}^{\frac{1}{2}},$$

where A is the sum of the correlations between test 'a' and the reference values 1, 2, 3; (i.e. correlations $r_{a1} + r_{a2} + r_{a3} + \ldots$) and A' is the sum of the squares of these correlations and T the sum of the intercorrelations between the reference values (i.e. $r_{12} + r_{13} + r_{14} \ldots$).

of Table III. Using these saturations, *residual* correlations are next calculated for Table II, with the results shown in Table IV. In Table IV the tests 1, 2, and 3 are omitted; all the residuals involving these tests are insignificant, their mean is zero (−0·00639), and each residual is

Table III

Test	Factor I (g)	Factor II (x)
1. G–1	0·652	—
2. G–2	0·756	—
3. G–3	0·612	—
4. Dearborn-Anderson	0·314	0·698
5. Cylinder Construction	0·340	0·476
6. Kohs's Block Design	0·618	0·559
7. Woolley	0·402	0·050
8. Passalong	0·297	0·062
9. Fitting Shapes	0·620	0·548
10. Shape Completion	0·393	0·549
11. Pool Image	0·415	0·589

Table IV. *Residual correlations of tests* 4–11 *after removal of g-factor*

	4	5	6	7	8	9	10	11
4		0·36109	0·35297	0·00688	0·05045	0·38185	0·41226	0·36960
5			0·23206	0·09373	0·09014	0·20006	0·31358	0·28567
6				0·07404	0·02543	0·33414	0·35142	0·32872
7					0·01302	−0·10648	0·10888	0·08267
8						0·02050	0·08911	−0·01791
9							0·24163	0·42206
10								0·26015
11								

itself insignificant as judged by the standard error of the difference between the original correlation and the theoretical one deriving from the g-factor content. The residuals for the remaining tests, however, are all positive (except $r_{8,11}$ and $r_{9,7}$). Assuming that one common factor (x) is sufficient to account for these residuals, we calculate the x-factor saturation of each test in turn in terms of all the other tests in the matrix, using Spearman formula (21) (1927). This gives the x-factor saturations shown in Table III, column (ii).

Inspection shows that the residuals after the removal of the x-factor are insignificant. The biggest of these second residuals is 0·09929; their mean is 0·00454. None is significant as judged by the standard error of the difference between the original correlation (first-factor residual) and the theoretical correlation deriving from the x-factor content.

(3) *Results of a multiple-factor analysis*

In addition to the above Spearman analysis a further examination of the data was made, using the Thurstone centroid method. In Table V

Table V. *Factor loadings for the eleven tests, using the Thurstone centroid method*

Test	Factor I	Factor II
1. G–1	0·472	−0·461
2. G–2	0·687	−0·418
3. G–3	0·553	−0·219
4. Dearborn-Anderson	0·629	0·389
5. Cylinder Construction	0·574	0·221
6. Kohs' Block Design	0·826	0·140
7. Woolley	0·411	−0·210
8. Passalong	0·340	−0·188
9. Fitting Shapes	0·756	0·309
10. Shape Completion	0·666	0·207
11. Pool Image	0·687	0·280

are shown the factor loadings obtained from this analysis, while Table VI (see p. 350) contains the residual correlations after removal of the two factors. These residuals are insignificant. The general resemblance between the loadings obtained by the second analysis and those of the Spearman analysis is very clear; rotation of axes would make it still more evident.

IV. Conclusion

Thus the two analyses, the Spearman and the Thurstone, both point to the same conclusion, viz. that the second factor in performance tests is the same as that in K-factor tests. K-factor of Koussi's study (1935) would seem to be the same as F-factor in Alexander's, as far as our data are concerned. It would seem best to retain the designation K to stand for the factor in either performance tests or in pencil-and-paper 'form-relations' tests, since what is apparently critical in both kinds of tests is the cognition and operation of 'form relations'. It is recognized, of course, that something intrinsic to performance might be critical for children; for our students, however, there would seem to be nothing critical in the performance as such.

The analysis makes it clear that there is not a group factor running through *all* performance tests, for the second factor saturations of the

Woolley and Passalong tests are negligible or non-existent. *K*-factor is not involved in the Passalong test, and it is not difficult to see why this should be so. The test is a performance test, it is true; but form and shape and spatial relations enter into it to a very slight extent. The test calls for that kind of thinking which is covered by *g*-factor. In theory, the

Table VI. *Residual correlations after removal of two factors (Thurstone centroid method)*

1	2	3	4	5	6	7	8	9	10	11
1	0·008	0·062	−0·016	0·135	−0·027	0·012	−0·131	−0·109	0·040	−0·007
2		−0·009	−0·016	−0·028	0·030	0·003	0·073	−0·074	0·013	−0·029
3			0·012	−0·074	−0·053	−0·104	−0·064	0·083	0·015	0·017
4				0·021	−0·027	0·044	−0·004	−0·020	−0·035	−0·041
5					−0·063	−0·040	−0·038	−0·091	0·019	−0·029
6						−0·031	0·045	0·050	0·015	−0·021
7							−0·048	0·103	−0·036	−0·027
8								−0·007	−0·019	0·076
9									−0·081	0·073
10										−0·093
11										

subject should be able to solve the later, more difficult items by reasoning upon his experience of the earlier ones. Thus the last two items present no difficulty to one who realizes that as they contain the same blocks as the last but one, differently arranged, the problem is merely to work back to the simpler arrangement. Actually this is not realized by the great majority of subjects; not six out of the eighty-five subjects tested declared that they could see any logical connexion between the various items. Nearly all declared that they had approached each one as a new problem, and, indeed, it was apparent from the random and fumbling approach of the great majority to each of the harder items, that they set about

them 'without an idea in their heads' of a logical kind. The later items in the Passalong test can be solved only by luck, or by explicit formulation of the 'trick' involved in the test. The Woolley test also failed to show correlation with a factor other than the general factor; the results obtained here throw no light on the relationship, if any, between haptic perception and *K*-factor.

REFERENCES

ALEXANDER, W. P. (1933). "A new performance test of intelligence." *Brit. J. Psychol.* XXIII, 52.

—— (1935). "Intelligence, concrete and abstract." *Brit. J. Psychol. Monog. Suppl.* XIX.

BROWN, W. & STEPHENSON, W. (1933). "A test of the theory of two factors." *Brit. J. Psychol.* XXIII, 352.

DEARBORN-ANDERSON TEST. *Reconstruction Puzzle or Formboard 2 of the Dearborn Series of Formboard and Performance Tests.* C. H. Stoelting Co. Catalogue No. 27164, p. 130.

HOLZINGER, K. (1937). *Student Manual of Factor Analysis.* University of Chicago.

KOHS, S. C. (1920). "The block-design tests." *J. exp. Psychol.* III, 357.

KOUSSI, A. A. H. (1935). "The visual perception of space." *Brit. J. Psychol. Monog. Suppl.* XX.

SMITH, I. (1937). "The form perception factor." M.Sc. Thesis, Glasgow University.

SPEARMAN, C. (1927). *The Abilities of Man.* Appendix, Formula (21). London: Macmillan.

SPEARMAN, C. & HOLZINGER, K. (1937). *Unitary Trait Study, Preliminary Report.* University of Chicago.

WOOLLEY TEST. *Identification of Forms Test.* C. H. Stoelting Co. Catalogue No. 36042, p. 164.

(*Manuscript received* 1 *December* 1938)

AN INVESTIGATION OF ABILITY IN WORK IN GROUPS AND IN ISOLATION

By N. P. MUKERJI

I. *Statement of the problem* (p. 352).
II. *Procedure* (pp. 352–353).
III. *Analysis of results* (pp. 353–354).
IV. *Oscillation* (pp. 354–356).
V. *Summary* (p. 356).
References (p. 356).

I. Statement of the problem

An effort has been made in the present article to examine if there be any difference in ability and in oscillation when a particular type of work is carried out by the same subjects working either in a group or in isolation.

II. Procedure

Thirty-one children, eighteen boys and thirteen girls, from an English elementary school[1] participated in the experiments; their ages ranged from 11 yr. 3 mo. to 13 yr. 8 mo. Two tests, viz. Cancellation and Naming Capitals, were first given to the subjects in mixed groups of not less than ten children. The tests were then repeated, but this time the subjects worked in isolation. In the Cancellation test they had to cancel out Ts and Xs only. In the Naming Capitals test, letters were typed in lines running in zigzags all over a sheet, and the subjects had to write S under the letters which consisted of straight lines only, e.g. A, K, L, M, and C under those which consisted wholly or in part of curved lines, e.g. O, P, Q, B.

The subjects were instructed to draw a line across the item they had reached whenever they heard a signal given at the end of each minute of work. Again, every 10 sec. after receiving a different signal they had to draw a line under the last item of work. Thus the whole test, which lasted for 5 min., was divided into thirty equal divisions of work.

Care was taken to choose only those subjects for the present experiments who had participated in the writer's previous experiments in which

[1] The author wishes to express his thanks to Prof. J. C. Flugel for his encouragement in this investigation and to the staff of the Isle of Dogs School, London, especially to Mr F. J. Shatwell, for their co-operation which made the study possible.

the tests mentioned above were included; this precaution was taken in order to minimize the effect of practice on the required data. Further, the subjects were given a half minute of practice trial before beginning work.

Each right answer scored one mark; the total number of marks obtained in 5 min. of work constituted the ability score. For the determination of reliability, the split-half method was used.

Henceforth, for the sake of abbreviation, the tests of Cancellation and Naming Capitals will be referred as T–X test and C–S test respectively.

III. ANALYSIS OF RESULTS

The reliability of the tests is shown in Table I. The correlations between the results of the two tests obtained in groups and in isolation is

Table I

	Test	
Groups	T–X	0·82 (0·039)
	C–S	0·77 (0·049)
Isolation	T–X	0·82 (0·039)
	C–S	0·87 (0·029)

(Figures within brackets denote P.E.)

shown in Table II; and the correlations between the two tests under these different conditions in Table III. Table IV shows the average

Table II

Test	
T–X	0·53 (0·091)
C–S	0·73 (0·059)

Table III

Groups	T–X and S–C tests	0·61 (0·076)
Isolation	Do.	0·42 (0·099)

Table IV

	Tests	Average ability	S.D.
Groups	T–X	149 (4·1)	34·2
	C–S	160 (5·3)	43·2
Isolation	T–X	114 (2·7)	23·0
	C–S	132 (2·8)	23·9

ability in the different tests. In the T–X test twenty-six subjects, i.e. 83·8 % of the cases, showed an increase in ability when they worked in groups. The average gain in the score (among the twenty-six subjects) was 27·7 %. In the case of the C–S test increased ability in group work was manifested by twenty-seven subjects, i.e. 87 %, and the average gain here was 20·0 %.

With regard to sex differences in ability the data collected here may not be highly significant as the number of subjects is comparatively small, but we may note that the difference is marked. It is shown in Table V

Table V

Pooled tests	Boys	Girls
In groups	153	156
In isolation	121	127

that the average ability among the boys is practically the same as for the girls, but there seems to be a probability of the girls scoring higher than the boys if N were larger. Taking the total raw scores in ability into consideration it is found on pooling the tests that the boys gained in group work by 61·5 % whereas the girls gained only 38·5 %.

IV. Oscillation

There are two chief possible ways of determining variability in oscillation. The first is by calculating the difference between the mean of the total work and the amount done in each short period of 10 sec. This procedure in itself is quite justifiable; but it can be argued that if the work gradually improves as the result of practice or of similar effects, then in adopting this method we shall be taking this improvement into account as well as oscillation. Since our concern is with oscillation and not with improvement, it is safer here to follow the other method which consists in determining the difference between successive individual measurements. This method has one advantage over the former: it measures short-period oscillation with a minimum of disturbance by any continuous improvement or deterioration of the work. It may be gathered from Table VI that oscillation as a function of ability remains

Table VI

	Tests	Average oscillation	S.D.
Groups	T–X	55·8 (1·4)	11·9
	C–S	56·3 (2·1)	17·8
Isolation	T–X	31·7 (0·6)	5·0
	C–S	31·7 (0·9)	8·2

practically unchanged whether tested in groups or in isolation. Comparing the total raw scores we find that the oscillation in group work is 28 % more than that in isolation. Thus, it may be stated at this stage that oscillation as a function of ability is positively related to it.

Table VII shows that oscillation as such does not vary extensively in the different sexes. The correlation between ability and absolute oscilla-

tion is shown in Table VIII. The correlations, though not uniformly high, are nevertheless all positive. Interpreting the above results, it may be said that the greater the ability, the greater the oscillation; but this effect is more marked in group work than in the work in isolation.

Another method of treating the oscillation data was tried by determining relative oscillation; the oscillation measurements already obtained

Table VII

Average oscillation among	In groups	In isolation
Boys	54	32
Girls	58	31

Table VIII

	Test	
Groups	T–X	0·51 (0·089)
	C–S	0·54 (0·085)
Isolation	T–X	0·21 (0·115)
	C–S	0·17 (0·117)

Table IX

	Tests	
Groups	T–X	−0·54 (0·085)
	C–S	−0·28 (0·109)
Isolation	T–X	−0·60 (0·077)
	C–S	−0·46 (0·095)

Table X

	Tests	
Groups	T–X	37·4
	C–S	36·5
Isolation	T–X	28·2
	C–S	24·3

Table XI

Average relative oscillation among	In groups	In isolation
Boys	36·5	27·9
Girls	37·5	25·2

(absolute oscillation) were calculated as percentages of the subject's total output. The results are shown in Table IX. We get here a result that contrasts strongly with the previous one; oscillation shows a negative relation to ability, and on the whole this relation is close. In general, the more able subjects show less oscillation, and this fact is corroborated by several previous workers (1,2,3).

The average relative oscillation in groups and in isolation is shown in Table X. Comparing these results with those in Table IV, we reach

conclusions which are in accordance with the correlations given in Table IX, namely, that average ability in the different tests shows an inverse relationship to average relative oscillation. Nevertheless, relative oscillation is greater in groups than in isolation. The results of the pooled tests, shown in Table XI, indicate that girls oscillate more when working in groups whereas boys oscillate more in isolation.

V. Summary

1. Thirty-one trained subjects (eighteen boys and thirteen girls) were given Cancellation and Naming Capitals tests in mixed groups of about ten. The tests were repeated on the same subjects, but this time they worked in isolation.

2. Ability in group work exceeded that in isolation by 20·0 % to 27·7 %. This advantage of group work is greater for boys than for girls, though the average score in ability is higher among the girls than the boys.

3. Absolute oscillation, between successive individual measurements, is positively related to ability. The oscillation in group work is greater than that in isolation, but the amount of oscillation is practically equal in boys and in girls.

4. Relative oscillation, determined in terms of the percentage of subject's total output, is negatively related to ability, i.e. the better subjects oscillate relatively less; however, relative oscillation is greater in group work than in isolation. Further, it is noticeable that relative oscillation is greater in the case of girls when working in groups whereas it tends to be less in boys under similar circumstances.

References

(1) Flugel, J. C. (1934). "Some recent studies of mental oscillation." *Ind. J. Psychol.* I.

(2) —— (1928). "Practice, fatigue and oscillation." *Brit. J. Psychol. Monogr. Supp.* No. XIII.

(3) Sengupta, N. & Sinha, C. (1934). "Mental work in isolation and in group." *Ind. J. Psychol.* I.

(4) Mukhopadhyay, N. P. (1938). "Oscillation at the threshold in relation to other mental functions." Ph.D. Thesis, in Library of University of London.

(*Manuscript received* 7 *November* 1939)

WEIGHTING FOR BATTERY RELIABILITY AND PREDICTION

By GODFREY H. THOMSON

Moray House, University of Edinburgh

I. *The Spearman-Brown formula* (pp. 357–358).
II. *Battery reliability* (pp. 358–359).
III. *Weighting for maximum battery reliability* (pp. 359–360).
IV. *A numerical example* (pp. 361–364).
V. *Reliability* versus *prediction of a criterion* (pp. 364–365).
VI. *Summary* (p. 365).
References (p. 366).

I. The Spearman-Brown formula

If we have a split test and find the correlation r of one half with the other we can, as is well known, estimate the correlation which the whole test will have with an alternative form of the whole by the formula

$$\frac{2r}{1+r},$$

which is therefore an estimate of the reliability of the whole test(1,2). The other whole test with which that which we possess is supposed to be correlated is of course imaginary. It is assumed that its halves would also correlate r with one another, and also that each of them would correlate r with each of the actual halves of the existing test. Also each of the four halves is assumed to have the same standard deviation. These assumptions can be clearly shown in a 'pooling square' (as I call it) by which Spearman's formula for the correlations of sums and differences can be readily applied. We have four half-tests h_1, h_2, h_1', h_2' (of which the last two are imaginary) and their intercorrelations:

	h_1	h_2	h_1'	h_2'
h_1	1	r	r	r
h_2	r	1	r	r
h_1'	r	r	1	r
h_2'	r	r	r	1

in which table only the r between h_1 and h_2 is an observed quantity, all the other r's being assumed. In such a pooling square (the standard deviations

being equal) the correlation between the battery h_1+h_2 and the battery $h_1'+h_2'$ is

$$\frac{\text{sum in N.E. quadrant}}{\sqrt{[(\text{sum in N.W.})\ (\text{sum in S.E.})]}}$$

$$=\frac{4r}{\sqrt{[(2+2r)\ (2+2r)]}}=\frac{2r}{1+r},$$

the Spearman-Brown formula. My object in writing this preliminary section is to prepare the way for the idea of the correlation of a *weighted* battery of tests with an imaginary similar and similarly weighted battery. For the Spearman-Brown formula envisages the simplest case of this idea, the (half-)tests in the battery being here alternative forms of the same (half-) tests, and the weights being equal.

II. Battery reliability

If in the above pooling square h_1 and h_2 were not halves of a test, but *different tests* in a battery of two tests, we would have the first approach to the more general case to be discussed later. We can then imagine h_1' to be an alternative form of h_1, and h_2' to be an alternative form of h_2. If we represent the correlation between h_1 and h_2 by r_{12}, we can, with the same justification as in § I, assume the correlation between h_1' and h_2' to be also r_{12}. Moreover the correlation between h_1 and h_2', and that between h_2 and h_1', can also be taken as r_{12}.

The correlation between h_1 and h_1', however, is now a different matter. It is the reliability of h_1, which we shall call r_{11}, and which we may assume to be estimated by splitting h_1 and 'stepping up' the split reliability by the Spearman-Brown formula. Similarly the correlation between h_2 and h_2' is r_{22}, the reliability of h_2. The pooling square is now

	h_1	h_2	h_1'	h_2'
h_1	1	r_{12}	r_{11}	r_{12}
h_2	r_{12}	1	r_{12}	r_{22}
h_1'	r_{11}	r_{12}	1	r_{12}
h_2'	r_{12}	r_{22}	r_{12}	1

and by the same principle as before the correlation of the battery h_1+h_2 with the (imaginary) battery $h_1'+h_2'$ is found (by summing the quadrants) to be

$$\frac{r_{11}+r_{22}+2r_{12}}{2+2r_{12}},$$

provided the standard deviations have all been equalized. This is the estimated reliability of the battery formed by the straight sum of its

constituent tests, calculated from the intercorrelations and separate reliabilities of those constituent tests.

If the reliability of the battery formed by a *weighted* sum of the constituent tests is required, the pooling square (which is of course only a way of envisaging Spearman's formula) gives the result almost equally easily. Each column *and each row* of the pooling square has to be multiplied by the appropriate weight before the quadrants are summed. Taking the weights to be w_1 and w_2 we have therefore, on thus multiplying the rows and columns, the square

	h_1	h_2	h_1'	h_2'
h_1	w_1^2	$w_1w_2r_{12}$	$w_1^2r_{11}$	$w_1w_2r_{12}$
h_2	$w_1w_2r_{12}$	w_2^2	$w_1w_2r_{12}$	$w_2^2r_{22}$
h_1'	$w_1^2r_{11}$	$w_1w_2r_{12}$	w_1^2	$w_1w_2r_{12}$
h_2'	$w_1w_2r_{12}$	$w_2^2r_{22}$	$w_1w_2r_{12}$	w_2^2

and the estimated reliability of the weighted battery is

$$\frac{w_1^2r_{11}+w_2^2r_{22}+2w_1w_2r_{12}}{w_1^2+w_2^2+2w_1w_2r_{12}}.$$

If the scores of the constituent tests have not been standardized, a σ with the appropriate subscript must accompany each w in this formula, for standard deviations act exactly like weights.

III. Weighting for maximum battery reliability

It is at once evident that in a given battery the estimated reliability of the weighted battery will depend not only on the test reliabilities and intercorrelations but on the weights chosen, and that there will exist a set of weights which make the estimated reliability a maximum. In the above simple case of two tests these 'weights for maximum battery reliability' can be found by elementary mathematical methods. But I shall consider the case of a larger battery and ask how then to find these weights. This is a special case of a more general problem solved by Hotelling(3). I shall first give the mathematical solution, and then a numerical example which will enable the non-mathematical reader to carry out the necessary calculations. To save space in printing I shall suppose the battery to contain three tests; but the method and formulae apply to any number.

Let the matrix of intercorrelations (including the test reliabilities) be as follows:

	z_1	z_2	z_3	z_1'	z_2'	z_3'
z_1	1	r_{12}	r_{13}	r_{11}	r_{12}	r_{13}
z_2	r_{12}	1	r_{23}	r_{12}	r_{22}	r_{23}
z_3	r_{13}	r_{23}	1	r_{13}	r_{23}	r_{33}
z_1'	r_{11}	r_{12}	r_{13}	1	r_{12}	r_{13}
z_2'	r_{12}	r_{22}	r_{23}	r_{12}	1	r_{23}
z_3'	r_{13}	r_{23}	r_{33}	r_{13}	r_{23}	1

$$= \begin{array}{c|c} A & C \\ \hline C & A \end{array}.$$

Here the correlations in the N.W. quadrant are experimental observations, and the reliabilities r_{11}, r_{22} and r_{33} are also supposed known, probably from splitting the tests z_1, z_2 and z_3 and 'stepping up' by the Spearman-Brown formula. The correlations r_{ij} in the N.E., S.W., and S.E. quadrants are assumptions, made with the same justification as the similar assumptions made in deducing the Spearman-Brown formula.

The capital letters A and C in the abbreviated version of the large matrix stand for the quadrant matrices. Thus

$$C = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{12} & r_{22} & r_{23} \\ r_{13} & r_{23} & r_{33} \end{bmatrix}$$

Then the method of finding the weights which give maximum reliability is to solve the equation

$$|CA^{-1}C - \lambda A| = 0$$

for its largest root λ_1; the weights in question are then any row of the adjugate of the matrix

$$(CA^{-1}C - \lambda_1 A),$$

and the maximum attainable reliability is the square root of λ_1.

These weights are to be applied to the constituent tests with equalized standard deviations. If the standard deviations are unequal, the weights must be divided, each by the corresponding standard deviation.

IV. A NUMERICAL EXAMPLE

Let the intercorrelations be as follows:

	z_1	z_2	z_3
z_1	1	0·482	0·617
z_2	0·482	1	0·397
z_3	0·617	0·397	1

and the estimated reliabilities 0·86, 0·73 and 0·83. (These data are taken from Brigham's *Study of Error*, 1932, p. 352.) The first step in the calculation is to find the triple matrix product $CA^{-1}C$. By far the best method of doing this is Aitken's(4), its advantages being more apparent when the number of tests is larger than in our example. A preliminary simplification can be carried out, since the matrices A and C differ only in their diagonal. If we write $D=A-C$, the diagonal matrix formed of the defects from unity of the reliabilities, i.e. in our case

$$\begin{bmatrix} 0{\cdot}14 & . & . \\ . & 0{\cdot}27 & . \\ . & . & 0{\cdot}17 \end{bmatrix}$$

it can easily be shown that

$$CA^{-1}C=A-2D+DA^{-1}D,$$

so that we have the easier task of calculating $DA^{-1}D$. This is done as follows. Write the three matrices in the form

A	D
$-D$	.

and condense by pivotal condensation till all the numbers are transferred into the south-east quadrant, which will then be the product-matrix $DA^{-1}D$. The process of pivotal condensation will be clear from the numerical example worked out in Table I. The top left-hand number (unity) is taken as 'pivot' and all the tetrad-differences of which it is a corner are set down in order. Thus 0·7677 is the value of $1-0{\cdot}482^2$, and 0·0996 is $0{\cdot}397-0{\cdot}482\times0{\cdot}617$. The whole matrix of six rows and six columns is thus reduced to one of five rows and five columns.

The same process is repeated, 0·7677 being taken as pivot. But before calculating the tetrad differences, since they all have to be divided by the pivot, it is convenient to divide the whole top row by 0·7677. Thus the row

	0·7677	0·0996	−0·0675	0·2700	. | 1·0698
becomes	1·0000	0·1297	−0·0879	0·3517	. | 1·3935.

This is written on the bottom edge of a slip of paper and fastened by two paper clips so as to replace the original row beginning with 0·7677. The pivot is now again unity, and the next slab of the calculation is formed by the tetrad-differences which have this pivot as one corner: thus 0·6064 is $0{\cdot}6193 - 0{\cdot}0996 \times 0{\cdot}1297$ and so on with the others.

Table I

1·000	0·482	0·617	0·140	.	.	2·239
0·482	1·000	0·397	.	0·270	.	2·149
0·617	0·397	1·000	.	.	0·170	2·184
−0·140	.	.	.	.	.	−0·140
.	−0·270	.	.	.	.	−0·270
.	.	−0·170	.	.	.	−0·170
	0·7677	0·0996	−0·0675	0·2700	.	1·0698
	0·0996	0·6193	−0·0864	.	0·1700	0·8025
	0·0657	0·0864	0·0196	.	.	0·1735
	−0·2700	.	.	.	.	−0·2700
	.	−0·1700	.	.	.	−0·1700
		0·6064	−0·0776	−0·0350	0·1700	0·6638
		0·0776	0·0255	−0·0237	.	0·0794
		0·0350	−0·0237	0·0950	.	0·1062
		−0·1700	.	.	.	−0·1700
			0·0354	−0·0192	−0·0218	−0·0056
			−0·0192	0·0970	−0·0098	0·0680
			−0·0218	−0·0098	0·0477	0·0161

Exactly the same process is repeated as often as necessary, until there are no numbers left in the left-hand block. The matrix $DA^{-1}D$ then appears in the middle block. The fact that it is symmetrical can either be used as a check, or to save about half the calculations all the way down, in which case, however, the signs have to be cared for. The check column is the sum of the row, and ought to agree in each row with the tetrad-difference formed from the appropriate rows in the preceding slab. If decimals are discarded the check column will show discrepancies in the last decimal kept, which should not exceed one or two places.

To form $CA^{-1}C$ we now have

$$CA^{-1}C = A - 2D + DA^{-1}D$$

$$= \begin{bmatrix} 1 & 0{\cdot}482 & 0{\cdot}617 \\ 0{\cdot}482 & 1 & 0{\cdot}397 \\ 0{\cdot}617 & 0{\cdot}397 & 1 \end{bmatrix} - \begin{bmatrix} 0{\cdot}280 & . & . \\ . & 0{\cdot}540 & . \\ . & . & 0{\cdot}340 \end{bmatrix}$$

$$+ \begin{bmatrix} 0{\cdot}035 & -0{\cdot}019 & -0{\cdot}022 \\ -0{\cdot}019 & 0{\cdot}097 & -0{\cdot}010 \\ -0{\cdot}022 & -0{\cdot}010 & 0{\cdot}048 \end{bmatrix} = \begin{bmatrix} 0{\cdot}755 & 0{\cdot}463 & 0{\cdot}595 \\ 0{\cdot}463 & 0{\cdot}557 & 0{\cdot}387 \\ 0{\cdot}595 & 0{\cdot}387 & 0{\cdot}708 \end{bmatrix}.$$

The equation for finding λ_1, the square of the maximum attainable reliability, is then

$$|CA^{-1}C-\lambda A|=0,$$

that is,

$$\begin{vmatrix} 0{\cdot}755-\lambda & 0{\cdot}463-0{\cdot}482\lambda & 0{\cdot}595-0{\cdot}617\lambda \\ 0{\cdot}463-0{\cdot}482\lambda & 0{\cdot}557-\lambda & 0{\cdot}387-0{\cdot}397\lambda \\ 0{\cdot}595-0{\cdot}617\lambda & 0{\cdot}387-0{\cdot}397\lambda & 0{\cdot}708-\lambda \end{vmatrix}=0.$$

This is a cubic, and the root required is the largest root. It is a great drawback that at this point an equation of the nth degree has to be solved, but luckily we always know approximately where the largest root lies, for it must be between the square of the largest single-test reliability, and unity: that is, in our case, between $0{\cdot}86^2$ and 1·00. In my experience, therefore, the easiest procedure at this stage is to evaluate the above determinant (by the same process[1] of pivotal condensation) for trial values of λ beginning with one a little larger than $0{\cdot}86^2=0{\cdot}7396$ and endeavouring to bracket the value which will make the determinant vanish. Thus if we try $\lambda=0{\cdot}8$ we have to evaluate the determinant

$$\begin{vmatrix} -0{\cdot}045 & 0{\cdot}077 & 0{\cdot}101 \\ 0{\cdot}077 & -0{\cdot}243 & 0{\cdot}069 \\ 0{\cdot}101 & 0{\cdot}069 & -0{\cdot}092 \end{vmatrix}=+0{\cdot}0033.$$

For $\lambda=0{\cdot}755$ the determinant $=+0{\cdot}0078$,
„ $\lambda=0{\cdot}9$ „ $=-0{\cdot}0086$,
„ $\lambda=0{\cdot}83$ „ $=+0{\cdot}0007$,
„ $\lambda=0{\cdot}836$ „ $=+0{\cdot}0001$.

We take $\lambda=0{\cdot}836$ as approximately the largest root, and

$$\sqrt{0{\cdot}836}=0{\cdot}914$$

as the maximum reliability attainable. The weights giving this will be proportional to any row of the adjugate of $(CA^{-1}C-0{\cdot}836A)$, i.e. of

$$\begin{bmatrix} -0{\cdot}081 & 0{\cdot}060 & 0{\cdot}079 \\ 0{\cdot}060 & -0{\cdot}279 & 0{\cdot}055 \\ 0{\cdot}079 & 0{\cdot}055 & -0{\cdot}128 \end{bmatrix},$$

that is to the minors got by crossing out any one row, and then each column in turn. We find the weights therefore to be approximately in the ratio

$$1 \quad 0{\cdot}36 \quad 0{\cdot}76,$$

[1] The determinant is condensed in the manner already shown (each pivot being converted into unity) until it is reduced to one number. The value of the determinant is then the product of all the *unconverted* pivots, including the last number arrived at. See Aitken (4).

We can, by a pooling square, check this result, by testing what reliability these weights would give, on the assumptions we are making. We set out the whole 6×6 matrix (though because of its symmetry half would serve), multiply rows and columns by these weights, and sum the quadrants:

	1	0·36	0·76	1	0·36	0·76
1	1·000	0·482	0·617	0·860	0·482	0·617
0·36	0·482	1·000	0·397	0·482	0·730	0·397
0·76	0·617	0·397	1·000	0·617	0·397	0·830
1	0·860	0·482	0·617	1·000	0·482	0·617
0·36	0·482	0·730	0·397	0·482	1·000	0·397
0·76	0·617	0·397	0·830	0·617	0·397	1·000

We obtain

3·209	2·936
2·936	3·209

and a reliability of

$$\frac{2 \cdot 936}{3 \cdot 209} = 0 \cdot 915$$

in agreement with our expectation ($\sqrt{0 \cdot 836} = 0 \cdot 914$).

V. Reliability *versus* prediction of a criterion

These weights for maximum battery reliability are of course different from the regression weights for the best prediction of an external criterion. Indeed the best reliability weights may give poor prediction, and the best prediction weights may give poor battery reliability. To illustrate this I choose a small and extreme example, admittedly not likely to occur in practice, but showing the principle involved with clearness. Let two tests z_1 and z_2, and an external criterion z_0, have the intercorrelations

	z_0	z_1	z_2
z_0	1·0	0·3	0·4
z_1	0·3	1·0	−0·4
z_2	0·4	−0·4	1·0

Also let the reliabilities of z_1 and z_2 be 0·5 and 0·8 respectively. With these data we can calculate the regression weights to give the best prediction of z_0 from a weighted battery of z_1 and z_2; and by a 4×4 pooling square we can find the reliability this weighted battery is likely to have.

On the other hand we can, by the method of §§ III and IV, find the best weights to give maximum reliability to a battery composed of z_1 and z_2;

and by a pooling square covering z_0, z_1 and z_2 we can find what correlation (i.e. prediction) these weights will give between the battery and z_0.

If, for convenience of comparison, we call the weight of z_2 unity in each case we find the following results:

Weights			
z_1	z_2	Prediction	Battery reliability
0·885	1·000	0·642 (max.)	0·450
−0·231	1·000	0·297	0·817 (max.)

We thus find that (in this example) the best prediction weights give an estimated battery reliability which is less than the multiple correlation, and less than the reliability of either test taken alone. On the other hand, the best weights for battery reliability give a prediction which is less than the correlation of either single test with the criterion. In this example weighting for prediction spoils reliability, and weighting for reliability spoils prediction. It seems natural to search for weights which will make battery reliability equal to prediction, and both as large as possible under this condition. In this simple case such weights can be found (though the process involves solving a quartic). They are weights of 0·5284 to 1·0000, and they give prediction equal to reliability, and both equal to 0·604. I hope that a young colleague to whom I am suggesting this problem may find a practicable way of calculating, for a battery larger than two, the weights which thus make the best of both prediction and battery reliability. I think also that the calculations here given for finding weights for maximum battery reliability will probably prove susceptible of simplification.

VI. Summary

The idea is mooted of an estimated battery reliability based on the reliabilities of the single tests and their weights in the battery. A method is given, with a numerical example, by which to calculate the weights which give the maximum battery reliability. This method is a special case of Hotelling's 'Most predictable criterion'. It is pointed out that prediction weights and reliability weights may conflict, and in a small example those weights are calculated which make prediction and reliability equal at the highest possible value under this condition.

REFERENCES

(1) SPEARMAN, C. (1910). "Correlation calculated from faulty data." *Brit. J. Psychol.* III, 290.

(2) BROWN, W. (1910). "Some experimental results in the correlation of mental abilities." *Brit. J. Psychol.* III, 299.

(3) HOTELLING, H. (1935). "The most predictable criterion." *J. educ. Psychol.* XXVI, 139–42.

(4) AITKEN, A. C. (1937). "The evaluation of a certain triple-product matrix." *Proc. roy. Soc. Edinb.* LVII, 172–81.

(*Manuscript received* 2 *December* 1938)